

**EXHIBIT T**

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF NEW YORK  
Civil Action No. 07-CV-2067

-----X  
THE UNITED STATES OF AMERICA,

Plaintiff,

and  
VULCAN SOCIETY, INC., for itself and on  
behalf of its members, CANDIDO NUNEZ,  
MARCUS HAYWOOD and on behalf of a Class  
of All Others Similarly Situated,  
Plaintiff-Intervenors

- against -

CITY OF NEW YORK, FIRE DEPARTMENT OF THE  
CITY OF NEW YORK, NEW YORK CITY  
DEPARTMENT OF CITYWIDE ADMINISTRATIVE  
SERVICES, MAYOR MICHAEL BLOOMBERG and  
NEW YORK CITY FIRE COMMISSIONER NICHOLAS  
SCOPETTA, in their Individual and  
Official capacities,

Defendants.  
-----X

271 Cadman Plaza East  
Brooklyn, New York

July 1, 2008  
9:39 a.m.

DEPOSITION of Non-Party Witness,  
PHILIP BOBKO, pursuant to Federal Rules of  
Civil Procedure, held at the above place,  
date and time, before Alice Schulman, a  
Notary Public of the State of New York.

VERITEXT REPORTING COMPANY

212-267-6868

516-608-2400

3935e4eb-2c85-4a5e-bf90-e41eddb0b8e4

1 P. Bobko

2 Bobko, Schemmer report that is, and we  
3 certainly used it, at least I did as  
4 context for the statement about job  
5 relatedness.

6 It's my understanding that the  
7 Morrongiello report refers to earlier work  
8 by Landy, Jacobs and whoever it is  
9 partner-wise. And so it formed some basis  
10 for the job relatedness statement.

11 Q. Did you do any comparison of the  
12 Landy, Jacobs test plan for exam 0084 to  
13 the Morrongiello test plan for exam 7029?

14 A. No.

15 Q. Is it correct that Landy, Jacobs  
16 never completed the job they planned  
17 because they didn't have the cooperation  
18 of the union?

19 A. I don't know.

20 Q. Do you know whether they ever  
21 completed the job analysis they planned?

22 A. No, I don't.

23 Q. What's your opinion of the  
24 transportability study that Landy, Jacobs  
25 did as reported in this document?

1 P. Bobko

2 A. I know there was a  
3 transportability study, or I believe there  
4 was some use of transportability from one  
5 city to another, but I haven't formed an  
6 opinion about that aspect of it. I wasn't  
7 asked to.

8 Q. So do you have an opinion as to  
9 whether it was a sufficient basis to  
10 establish the validity of the written and  
11 physical exam established by Landy for New  
12 York?

13 A. I don't have an opinion.

14 Q. Back in your report, Exhibit  
15 601, the next document listed is  
16 Probationary Firefighter's Manual, Fire  
17 Department of the City of New York.

18 For what purpose did you rely on  
19 this document in forming your opinions and  
20 conclusions?

21 A. Two purposes. One, it was just  
22 a quick scan was a helpful way to  
23 understand yet again the job of  
24 firefighter and get a sense of the context  
25 of the selection process.

1 P. Bobko

2 Q. So you talked about those things  
3 as positives and negatives, is that what  
4 you're saying?

5 A. I don't remember who said what.  
6 I know those things were discussed, and I  
7 know I was at that meeting or those series  
8 of meetings.

9 Q. Let me ask this then: Did you  
10 provide any input or did you just sit back  
11 and listen to the discussion going on?

12 A. I'm sure I said something. I  
13 don't know what it was.

14 Q. Do you recall any discussion  
15 about the possibility of using the CPAT  
16 competitively, in other words, not scoring  
17 it pass/fail, but scoring it in a way that  
18 would allow the City to combine the score  
19 on the CPAT, if we can call it that, and  
20 the written exam?

21 A. I don't recall any specific  
22 discussion about how one might score the  
23 CPAT other than dichotomously.

24 Q. What is the relative importance  
25 of physical abilities, cognitive abilities

1 P. Bobko

2 and personal attributes to the successful  
3 performance of the FDNY entry level  
4 firefighter job?

5 A. I don't know. I can tell you  
6 they are all important based on the job  
7 analysis, but that's all I can tell you.

8 Q. When you say "based on the job  
9 analysis," what job analysis are you  
10 referring to?

11 A. Well again, that would be the  
12 Morrongiello job analysis, the job  
13 analysis that was done for 6019, and I  
14 guess to some extent the job analysis that  
15 was done for what I'll call the Landy exam  
16 0084 whatever that number is.

17 Q. 0084.

18 Let me show you what's been  
19 marked as Exhibit 242, and for the record  
20 this is a document that is Bates numbered  
21 DCAS 0016940 to DCAS 0016945, and on the  
22 top of the first page it says Physical and  
23 Cognitive Contributions to Task Area  
24 Performance. Dr. Bobko, have you ever  
25 seen this document before?

1 P. Bobko

2 a minute to check my report?

3 Q. Please go ahead.

4 A. Thank you. Could you please  
5 state the question again, but I think I'm  
6 getting there.

7 Q. My question was: In your  
8 opinion, did the City's use of exam 7029  
9 as a pass/fail screening device with a  
10 cutoff score of 84.705 result in a  
11 disparate impact on black candidates?

12 A. Yes.

13 Q. In your opinion, did the City's  
14 use of written exam 7029 as a pass/fail  
15 screening device with a cutoff score of  
16 84.705 result in a disparate impact upon  
17 Hispanic candidates?

18 A. No.

19 Q. In your opinion, did the City's  
20 use of written exam 2043 as a pass/fail  
21 screening device with a cutoff score of 70  
22 result in a disparate impact upon black  
23 candidates?

24 A. If I may, you asked me to try to  
25 answer your questions yes and no. I am

1 P. Bobko

2 trying to do that.

3 Q. Okay.

4 A. So on the Hispanics when I said  
5 no, I know there are, understand there are  
6 people who would claim otherwise. I'm  
7 trying to give you the best answer I can  
8 between yes and no. Is that okay?

9 Q. Let me clarify then just to make  
10 sure I don't misinterpret your answers.

11 A. Sure.

12 Q. If you have an opinion that is  
13 yes or no, it either does or doesn't, in  
14 your opinion have a disparate impact on  
15 whatever group I'm asking you, then yes,  
16 please answer yes or no.

17 If you don't have an opinion  
18 that it does and you don't have an opinion  
19 that it doesn't have an opinion that it  
20 has a disparate impact, please tell me  
21 that, and I'm not asking you for other  
22 opinions others may have at this point.  
23 Does that help?

24 A. No, I have an opinion about what  
25 opinion others have.



1 P. Bobko

2 MR. FRAENKEL: If your opinion  
3 is not yes or no but is something  
4 other than that, give your opinion as  
5 to what the other is. You do not have  
6 to answer yes or no unless that answer  
7 is your answer.

8 Q. Tell me if you can't answer yes  
9 or no, and I'll follow up and try to  
10 figure out.

11 A. I can't answer yes or no on the  
12 answer on Hispanic.

13 Q. Would you need some additional  
14 information to be able to answer yes or  
15 no?

16 A. I would need some additional  
17 clarification.

18 Q. What clarification would you  
19 need?

20 A. How one defines adverse impact.

21 Q. Okay, then let's go to the next  
22 question now that we cleared that up.

23 In your opinion, did the City's  
24 use of exam 2043 as a pass/fail screening  
25 device with a cutoff score of 70 result in

1 P. Bobko

2 a disparate impact upon black candidates?

3 A. I can't answer that question  
4 either.

5 Q. And is that for the same reason  
6 as with regard to the previous question?

7 A. Yes.

8 Q. In your opinion, did the City's  
9 use of written exam 2043 as a pass/fail  
10 screening device with a cutoff score of 70  
11 result in a disparate impact upon Hispanic  
12 candidates?

13 A. I can't answer that question  
14 either.

15 Q. And, again, is that because  
16 you'd have to know how one defines  
17 disparate impact?

18 A. Correct. I would not need to  
19 know how one defines it, I would need to  
20 know how you define it or the person  
21 asking the question would define it.

22 Q. Is there any other information  
23 you would need to know in order to give a  
24 yes or no answer?

25 A. No.

1 P. Bobko

2 Q. In your opinion, did the City's  
3 rank/order/processing and selection from  
4 the exam 7029 eligibility list result in a  
5 disparate impact upon black candidates?

6 A. Could you repeat the question,  
7 please.

8 Q. In your opinion, did the City's  
9 rank/order/processing and selection from  
10 the exam 7029 eligibility list result in a  
11 disparate impact upon black candidates?

12 A. Generally, yes.

13 Q. What do you mean when you say  
14 "generally, yes"?

15 A. Generally, is because I don't  
16 have in front of me the exact, I do, but I  
17 don't recall the exact numbers that  
18 plaintiffs' expert's report provide for  
19 the situation you outlined, and I think  
20 there is a leap in your question between  
21 having an adverse impact, once again, I'm  
22 not quite sure I understand fully the  
23 concept of adverse impact as defined in  
24 your situation of rank/order/processing.

25 Q. Okay. Let me ask this, I may

1 P. Bobko

2 well get the same answer. In your opinion,  
3 did the City's rank/order/processing and  
4 selection from the exam 7029 eligibility  
5 list result in a disparate impact on  
6 Hispanic candidates?

7 A. That one I can't answer.

8 Q. Why can't you answer?

9 A. For the same reason when you ask  
10 about the pass/fail.

11 Q. In your opinion, did the City's  
12 rank/order/processing and selection from  
13 the exam 2043 eligibility list result in a  
14 disparate impact upon black candidates?

15 A. I can't answer.

16 Q. Is that for the same reason as  
17 you can't answer the previous question?

18 A. Correct.

19 Q. In your opinion, did the City's  
20 rank/order/processing and selection from  
21 the exam 2043 eligibility list result in a  
22 disparate impact upon Hispanic candidates?

23 A. I can't answer it.

24 Q. For the same reason?

25 A. Same reason.

1 P. Bobko

2 adverse impact, but the practice really  
3 does not have adverse impact; is that  
4 right?

5 A. Correct.

6 Q. I'm handing you what's been  
7 marked as Exhibit 614, and this document  
8 at the top has a title Modeling the  
9 Behavior of the 4/5ths Rule for  
10 Determining Adverse Impact: Reasons for  
11 Caution. Do you recognize this document?

12 A. Yes.

13 Q. Exhibit 614 is a copy of an  
14 article that was written by you and two  
15 others published in 2006 in the Journal of  
16 Applied Psychology, correct?

17 A. Correct.

18 Q. In your 2006 article, Exhibit  
19 614, you talk about false positives and  
20 false negatives. And if you want to take  
21 a look, an example of that is on page 509  
22 the top left column.

23 Do you see what I'm talking  
24 about in that first paragraph on the page?

25 A. Yes.

1 P. Bobko

2 Q. And what I want to do is just  
3 clarify for myself what a false positive  
4 and a false negative is as you use the  
5 terms in this article. I know they may  
6 have other meanings elsewhere.

7 What you call a false negative  
8 in this article is the situation where a  
9 practice does have an adverse impact, but  
10 the 80 percent rule does not indicate  
11 adverse impact; is that right?

12 A. Correct.

13 Q. And the situation in which a  
14 practice does not have an adverse impact  
15 but the 80 percent rule indicates that it  
16 does is what you call a false positive in  
17 this article, correct?

18 A. What Greenberg called a false  
19 positive.

20 Q. And you use the term throughout?

21 A. That's correct.

22 Q. And it has the same meaning  
23 throughout the article?

24 A. Correct.

25 Q. Do you agree that a test of

1 P. Bobko

2 alpha, A-L-P-H-A, level of .05 or .01. I  
3 don't read any link to controlling false  
4 positives in that sentence.

5 Q. That particular sentence?

6 A. Yes.

7 Q. Type I errors, those are false  
8 positives, right?

9 A. Correct.

10 Q. And what you call false  
11 negatives in your 2006 article, Exhibit  
12 614, that's also what you sometimes call  
13 Type II errors, correct?

14 A. Correct.

15 Q. In your opinion, if you have a  
16 large sample, which is better at  
17 controlling false negatives or Type II  
18 errors, a statistical significance test or  
19 the 80 percent rule of thumb?

20 A. Just bear with me because of the  
21 Type I and Type II. Your question was  
22 about Type II, correct?

23 Q. Right.

24 A. Can you repeat your question,  
25 please?

1 P. Bobko

2 Q. Sure. My question was: In your  
3 opinion, if you have a large sample, which  
4 is better at controlling false negatives,  
5 a statistical significance test or the 80  
6 percent rule of thumb.

7 A. One more time, repeat the  
8 question and then I'll give you your  
9 answer.

10 Q. In your opinion, if you have a  
11 large sample, which is better at  
12 controlling false negatives, a statistical  
13 significance test or the 80 percent rule?

14 A. In general, a statistical  
15 significance test with large samples.

16 Q. Do you agree with this  
17 statement, with a statistical significance  
18 test all elements equal, the power to  
19 detect differences in selection rates and  
20 avoid Type II errors increases with  
21 increasing sample size?

22 A. That's correct.

23 Q. If you would look at Exhibit  
24 614, page 514. Are you there?

25 A. Yes.



1 P. Bobko

2 Q. On page 19 at the bottom you  
3 have some bullet points carried over to  
4 the next page that state the basis for  
5 your opinion that exactly no difference is  
6 an unrealistic standard here, right?

7 A. Correct.

8 Q. So the literature and other  
9 things you mention in these bullets, as  
10 you said, lead you to expect that blacks  
11 and Hispanics will score lower than whites  
12 on traditional paper and pencil cognitive  
13 tests, right?

14 MR. FRAENKEL: Objection as to  
15 form. You can answer.

16 A. I disagree.

17 Q. Why?

18 A. Because you use the word paper  
19 and pencil.

20 Q. I wasn't limiting it to paper  
21 and pencil. I was saying if it is a paper  
22 and pencil test.

23 A. But so which question do you  
24 want me to answer?

25 Q. I want you to answer this

1 P. Bobko

2 question, the literature and the other  
3 things that you list in the bullets here  
4 lead you to expect that blacks and  
5 Hispanics will score lower than whites on  
6 a traditional paper and pencil cognitive  
7 test; is that correct?

8 MR. FRAENKEL: Objection as to  
9 form, but you can answer.

10 A. That's correct.

11 Q. And both the written exams at  
12 issue in this case, written exam 7029 and  
13 written exam 2043 are paper and pencil  
14 tests that, in your opinion, each assess  
15 cognitive ability, right?

16 A. Yes, fashion sets of cognitive  
17 ability, yes.

18 Q. So you would expect that black  
19 FDNY firefighter candidates would score  
20 lower on written exam 7029 than white  
21 candidates, right?

22 A. On average, yes.

23 Q. And would you expect that black  
24 candidates would score lower on written  
25 exam 2043 on average than white

1 P. Bobko

2 been hired. These estimates are  
3 overstated and conceptually flawed."

4 So my first question is this  
5 criticism, this section is a criticism of  
6 Dr. Siskin's and Dr. Wiesen's statements  
7 of hiring shortfalls, right?

8 A. Yes.

9 Q. There's nowhere in your report  
10 where you criticize Dr. Siskin's  
11 calculations of the shortfall in test  
12 passers, correct?

13 A. There's no -- could you say your  
14 sentence again, please, the question.

15 Q. My question is, is it correct  
16 that there's nowhere in your report where  
17 you criticize Dr. Siskin's calculation of  
18 the shortfall in test passers as opposed  
19 to hiring shortfalls?

20 A. That's correct.

21 Q. I want to go basically through  
22 your specific complaints or criticisms  
23 that you do make in your report and  
24 beginning on page 20, the first one is  
25 that you state that, in your opinion,

1 P. Bobko

2 shortfalls have not been discussed in any  
3 type of case law?

4 A. No.

5 Q. Is that because you're not  
6 familiar with Title VII case law?

7 A. Correct.

8 Q. Your next complaint about Dr.  
9 Siskin's shortfall estimates has to do  
10 with variability, right?

11 A. Yes.

12 Q. And your complaint is that the  
13 estimated hiring shortfalls have  
14 associated margins of error that Dr.  
15 Siskin didn't report; is that right?

16 A. That's what we are pointing out,  
17 yes.

18 Q. So as you say in the last  
19 sentence of this section labeled  
20 variability on page 21, although the  
21 underlying, actual shortfalls might, in  
22 fact, be higher, they might also be lower,  
23 right?

24 A. Correct.

25 Q. So your criticism essentially is

1 P. Bobko

2 that Dr. Siskin didn't give us in his  
3 report any probabilities related to how  
4 much lower or higher the actual shortfalls  
5 might have been; is that right?

6 A. Correct.

7 Q. But the probabilities would be  
8 symmetrical, wouldn't they?

9 A. Technically I could imagine that  
10 they may not be, but my guess is  
11 practically speaking they would be very  
12 similar and therefore functionally  
13 symmetrical.

14 Q. If you had to give a point value  
15 estimate, Dr. Siskin's is in the middle,  
16 it's the best estimate, right?

17 A. To you.

18 Q. Is it not the best estimate if  
19 you had to give a point value estimate?

20 A. I can't answer that question yes  
21 or no.

22 Q. Why not?

23 A. It depends on what you define as  
24 best.

25 Q. Okay. If you had to give a

1 P. Bobko

2 remember, also you don't want to estimate  
3 things by using the middle.

4 Q. So you didn't state any opinion  
5 that Dr. Siskin's use of the middle, as  
6 we've been calling it, is not the best  
7 estimate in this situation, correct?

8 A. Correct.

9 Q. Now, you agree, don't you, that  
10 we can't know what you call the actual  
11 shortfalls because people who failed or  
12 scored too low on the City's test weren't  
13 allowed to go on and process, right?

14 A. That makes sense, I don't  
15 remember saying that.

16 Q. But it's true, correct?

17 A. It sounds correct to me, yes.

18 Q. Is there something that leads  
19 you to be unsure as to whether that's  
20 uncorrect or not?

21 A. No.

22 Q. And then still on page 21, your  
23 last criticism on this page about the  
24 hiring shortfalls is as a measure of  
25 practical significance is they ignore

1 P. Bobko

2 A. Correct.

3 Q. Your next criticism of Dr.  
4 Siskin's delay analysis is that the delay  
5 statistics in Dr. Siskin's report is  
6 singled value estimates that are  
7 associated with margins of error that are  
8 not reported, correct?

9 A. Correct.

10 Q. So essentially is this the same  
11 complaint you stated about Dr. Siskin's  
12 shortfalls?

13 A. It's the same type of complaint,  
14 yes.

15 Q. Now, just to make sure I  
16 understand what you mean with the  
17 complaint in this context, I wanted to ask  
18 you to look at Exhibit 611 which was Dr.  
19 Siskin's report, and go back to the tables  
20 which are at the end of the report. Well,  
21 they're before all the appendices, so it  
22 will be more like the middle of the  
23 report. I can give you a Bates number,  
24 USAEXP00039, which is table 3, Part A, and  
25 then it's followed by table 3, Part B, and

1 P. Bobko

2 these tables deal with his delay  
3 statistics for blacks for exam 7029.

4 First, I want to look at table  
5 3, Part A and in particular the column  
6 labeled Actual. The numbers in the Actual  
7 column are not estimates; is that correct?

8 A. That's my understanding,  
9 correct.

10 Q. Looking at the column labeled  
11 Expected. Are the numbers in the Expected  
12 column the statistics that you are  
13 referring to in your report?

14 A. Yes.

15 Q. To make sure I'm clear, those  
16 numbers are the results of a calculation  
17 of how many blacks would have been  
18 appointed in each class if the  
19 distribution of appointment dates were the  
20 same for blacks as for whites; is that  
21 your understanding?

22 A. Yes.

23 Q. And the numbers in the shortfall  
24 column in Table 3, Part A are just what  
25 you would get by subtracting the numbers



1 P. Bobko

2 in the Actual column from the numbers in  
3 the Expected column, right?

4 A. Yes.

5 Q. So that's not another estimate,  
6 that's just arithmetic?

7 A. Correct.

8 Q. So exactly what is your  
9 criticism as it relates to Table 3A?

10 A. The criticism in respect to what  
11 page on our report, please.

12 Q. Page 22, I'm sorry, it's 23 in  
13 your report and it's the subheading  
14 Variability.

15 A. The criticism about no  
16 variability, the criticism about no  
17 variability is exactly that, that as you  
18 point out, the second and, the last column  
19 and next to last column are statistically  
20 calculated values from Dr. Siskin's report  
21 because he calculates the expected values  
22 and takes those numbers and changes them  
23 into collated arithmetic shortfall  
24 values. Because they are statistical  
25 estimates, they do, or they can be

1 P. Bobko

2 associated with variability sampling,  
3 variability as is typically the case in  
4 any statistical hypothesis. And yet they  
5 are reported here in Table 3 as single  
6 valued numbers without any indication of  
7 how much variability there is around  
8 those.

9 So the first number, shortfall  
10 number is three. Who knows what the  
11 variability, statistical variability  
12 around that number 3 is, it could be  
13 large, it could be small. That is what we  
14 were trying to say in the first paragraph  
15 of the variability statement.

16 Q. So it's the same conclusion,  
17 that Dr. Siskin gives you estimates, but  
18 the actual numbers could be higher, could  
19 be lower?

20 A. Correct.

21 Q. Let's look at Table 3B just to  
22 make sure I am completely clear. Now,  
23 these numbers, Table 3B of Dr. Siskin's  
24 report which is the Bates number  
25 USAEXP00040, the numbers on this table are

1 P. Bobko

2 job relatedness and business necessity?

3 MR. FRAENKEL: Objection to  
4 form. You can answer.

5 A. Let me make sure I get the  
6 question right because it had a long  
7 preface. You're asking if the  
8 documentation in the Bobko, Schemmer  
9 report is sufficient to establish job  
10 relatedness?

11 Q. Let me say it again.

12 A. Please.

13 Q. Is what's in this report, the  
14 Bobko, Schemmer report, sufficient to  
15 establish that the City's use of exam 7029  
16 as a pass/fail screening device with a  
17 cutoff point of 84.705 is consistent with  
18 job relatedness and business necessity?

19 A. No.

20 Q. Is it your opinion it is  
21 sufficient to establish that the City's  
22 use of written exam 2043 as a pass/fail  
23 screening device with a cutoff score of 70  
24 is job related and consistent with  
25 business necessity?

1 P. Bobko

2 A. No.

3 Q. Is it your opinion that what is  
4 in this report is sufficient to establish  
5 that the City's rank/order/processing and  
6 selection of candidates from the exam 7029  
7 eligibility list is job related and  
8 consistent with business necessity?

9 A. No.

10 Q. Is it your opinion that what is  
11 in this report is sufficient to establish  
12 that the City's rank/order/processing and  
13 selection of candidates from the exam 2043  
14 eligibility list is job related and  
15 consistent with business necessity?

16 A. No.

17 Q. Why did you focus so much on  
18 disparate impact in your report and so  
19 little on job relatedness and business  
20 necessity?

21 A. Why is a difficult question to  
22 ask -- answer. I don't know completely  
23 why. I know that when we started writing  
24 this report, we thought, at least I  
25 thought the purpose of the report was a

1 P. Bobko

2 that you mean their answers to the items?

3 A. Their answers to the items.

4 Q. Go on.

5 A. They were responding an as if  
6 they were or they were applicants. And  
7 I've now lost track of your question.

8 Q. The question was, now I know  
9 what data you were talking about. What  
10 would you be correlating with what?

11 A. I would be correlating scores,  
12 people's scores on each item with their  
13 scores on other items to find what's the  
14 relationship between item 17 and item 22,  
15 what's the relationship between item 17  
16 and item 23, et cetera.

17 Q. And what would you be looking  
18 for, particular pattern of correlations or  
19 what?

20 A. That's a tough one and that's  
21 why there's nothing definitive in the  
22 suggestion that I offered to you as what I  
23 might do. It would be useful if all of  
24 the math questions, well, all the written  
25 comprehension questions, let's say the

P. Bobko

first ability listed there, if they all tended to correlate more highly among themselves than they correlated with items measuring memorization. That would give me some further evidence that oh, yeah, this was a reasonable set of questions having to do with written comprehension because every time I came up with a question it seemed to correlate pretty highly with another question.

The caveat being, as you pointed out this morning, this is this notion of gee, it's hard to discern those kinds of patterns because very often the subfacets are correlated very highly with the other subfacets, so it's hard to tease all of it out.

Q. You didn't do an analysis of what you were just talking about; is that right?

A. Correct.

Q. But the data was available to do that, correct?

A. Yes.

1 P. Bobko

2 there was more than one form of validity  
3 generalization, you thought the answer was  
4 yes. Why is that?

5 A. My guess is if we had ten  
6 researchers in a room and we asked them to  
7 do a validity generalization statement,  
8 they would all do something different.

9 Q. Let me ask this, do you agree  
10 that what Dr. Landy did in doing a  
11 transportability study to determine  
12 whether the job of firefighter in D.C. was  
13 similar enough to a job of firefighter in  
14 New York is something different than what  
15 you're relying on in this case which does  
16 not involve any transportability study?

17 MR. FRAENKEL: Objection as to  
18 form.

19 A. I can't begin to answer your  
20 question.

21 Q. Let me ask a simpler question.  
22 In your report, you don't rely on any  
23 transportability study, do you?

24 A. No.

25 Q. In your report, you rely on

1 P. Bobko

2 validity generalization based on  
3 meta-analysis, right?

4 A. Only in an extremely limited  
5 sense.

6 Q. What do you mean by that?

7 A. The only place I recall in our  
8 report that we were relying on validity  
9 generalization is to point out that, well,  
10 actually two places. One, that cognitive  
11 ability tests generally have validity  
12 across most jobs and two, when we cite, I  
13 think it's Barrett, Barrett et al 1999  
14 paper in footnote 35.

15 Q. Well, you don't cite any other  
16 evidence of criterion-related validity for  
17 firefighter exams, right?

18 A. Correct.

19 Q. The only criterion-related  
20 evidence you talk about for firefighters  
21 is that Barrett article, right?

22 A. Correct.

23 Q. And that's the article that's  
24 cited in footnote 35 of your report,  
25 correct?



1 P. Bobko

2 A. Yes.

3 Q. I'm handing you what's been  
4 marked Exhibit 620. Dr. Bobko, is this  
5 the Barrett meta-analysis that you cite in  
6 your report?

7 A. I believe so, yes.

8 Q. This article was published in  
9 the Journal of Business and Psychology,  
10 correct?

11 A. Correct.

12 Q. Are you familiar with that  
13 journal?

14 A. I know it exists.

15 Q. Is it a peer review journal?

16 A. I don't know.

17 Q. Have you published in that  
18 journal?

19 A. I don't think so.

20 Q. Do you know the quality or  
21 reputation of that journal?

22 A. No.

23 Q. Now, the Barrett article was  
24 published in 1999, right?

25 A. Correct.

1 P. Bobko

2 Q. So when the authors say, I  
3 believe it's on page 509 of this article,  
4 yes, it's the first sentence. "The data  
5 used in this study was compiled from  
6 criterion-related studies of cognitive  
7 ability and mechanical/spatial tests  
8 conducted over the past two decades." Do  
9 you understand the past two decades to  
10 mean from 1979 to 1999?

11 A. Yes.

12 Q. Did you review any of the  
13 studies used by Dr. Barrett for  
14 meta-analysis?

15 A. I don't think so.

16 Q. Do you know what measure of job  
17 performance was used as criterion in any  
18 of the studies of the meta-analysis?

19 A. Measures of job performance?

20 Q. Yes.

21 A. No.

22 Q. Now --

23 A. Other than he clearly says, or  
24 they clearly say criteria consisted of  
25 supervisory ratings, so I assume it

1 P. Bobko

2 consisted of repeat.

3 Q. That's all you know?

4 A. That's all I know.

5 Q. On your report on page 27, why  
6 do you say there were more than 100  
7 samples in the Barrett study?

8 A. I'm looking at the same page 509  
9 of the Barrett et al cite, the second  
10 paragraph. He, they talk about 73 samples  
11 for that, this performance criteria and 28  
12 for training. I'm wondering if we added  
13 those two together, 73 and 28 and got a  
14 number greater than 100. Although, as I  
15 sit here today, I don't think I could tell  
16 you whether there might have been some  
17 overlap between the 73 samples and the 23  
18 samples.

19 Q. Would you look at the last page  
20 of this Exhibit 620 and see Appendix B at  
21 the bottom of that page. Do you see that?

22 A. Yes.

23 Q. How many samples does this  
24 indicate the authors used?

25 A. 25. I count 25.

1 P. Bobko

2 Q. So it wasn't over 100, right?

3 A. 25 is certainly less than 100.

4 Q. But there are more r's used than  
5 that according to this?

6 A. Yes.

7 Q. For the record, what is r in  
8 this context?

9 A. The correlation coefficient  
10 between the test score and the performance  
11 measure.

12 Q. If we look at page 510 of this  
13 exhibit, and if we count up the r's in  
14 Table 1 and Table 2, in Table 1 there are  
15 73 r's and in Table 2 there are 28, right?

16 A. Yes.

17 Q. So those numbers that they were  
18 adding up, those were r's, not samples,  
19 right?

20 A. Apparently.

21 Q. If we look again at Appendix B,  
22 the second column that's labeled N of r's,  
23 when you add up all the r's there, there  
24 are 105 of them and you can either add  
25 them or take my word for it. Do you know

1 P. Bobko

2 A. I don't think so.

3 Q. Are you at all familiar with any  
4 of the studies that were used in this  
5 meta-analysis?

6 A. No.

7 Q. Now, the studies listed are from  
8 between 1958 and 1986; is that right?

9 A. I'm sure you're right. I'm just  
10 perusing, I didn't see the 1950's one, but  
11 it's probably accurate.

12 Q. If that's accurate, then the  
13 authors present, right, when they said the  
14 studies were from the past two decades,  
15 right?

16 A. I don't know if most of the  
17 studies were from the past two decades.

18 Q. Does it say that most of the  
19 studies were from the past two decades?

20 A. No.

21 Q. In fact, it looks like the date  
22 of the Barrett meta-analysis in 1999, less  
23 than half of the studies were within the  
24 past 20 years and now less than none of  
25 them are less than 20 years old?

1 P. Bobko

2 A. Right.

3 Q. You don't know anything other  
4 than the dates listed in here?

5 A. That's correct.

6 Q. Did the Barrett meta-analysis  
7 include all the criterion-related studies  
8 for cognitive firefighter exams that had  
9 been conducted prior to 1999 by Dr.  
10 Barrett or either of his co-authors?

11 A. Let me make sure I follow your  
12 question. You're asking me does this  
13 meta-analysis include studies by Gerry  
14 Barrett?

15 Q. Gerry Barrett or either of his  
16 two co-authors had done up to that point  
17 in time?

18 A. I have no clue.

19 Q. Do you know of any that were not  
20 included?

21 A. No.

22 Q. Were there other  
23 criterion-related studies for firefighter  
24 tests for firefighters that were available  
25 in 1999 but weren't included in this

1 P. Bobko

2 the number of r's is seven. It could be  
3 that there were subsamples in those two  
4 samples and that each of those subsamples  
5 included different people, and so the  
6 seven r's come from truly seven  
7 independent subsamples.

8 Q. But you don't know that, right?

9 A. But I don't know that, but it's  
10 why I can't agree with your question.

11 Q. Let me ask it this way. We  
12 don't know from this article how many  
13 independent samples there were; is that  
14 right?

15 A. We know what they say, that's  
16 all we know.

17 Q. We don't know how many  
18 independent samples there were?

19 A. I can't answer that question.

20 Q. Do you know how many independent  
21 samples there were?

22 A. No.

23 Q. Now, if you look back at  
24 Appendix B, for one of the studies the  
25 authors say, and I believe this is the

1 P. Bobko

2 Kriska & Hines, there were 19 r's from one  
3 sample, right, do you see that?

4 A. Yes, I do.

5 Q. So for all we know, the 24 r's  
6 listed in Table 1 for cognitive tests all  
7 came from no more than a handful of  
8 independent samples, right?

9 A. Again, I can't answer that  
10 question if there were independent  
11 subsamples.

12 Q. Does it say that anywhere in the  
13 report?

14 A. It doesn't say that anywhere.

15 Q. The issue of independent samples  
16 is pretty important for meta-analysis,  
17 right?

18 A. I think so.

19 Q. Now, you agree, don't you, that  
20 the formulas used to compute percent  
21 sampling variance, in effect error  
22 confidence intervals, et cetera, typically  
23 assume that samples are independent?

24 A. It's my understanding.

25 Q. Do you agree that it is very



1 P. Bobko

2 Q. Let me make sure of one thing.  
3 In your report, you didn't cite any other  
4 criterion study on firefighters; is that  
5 right?

6 A. That's correct.

7 Q. Now, back in your report on page  
8 27, you say that Barrett and his  
9 co-authors found that cognitive and  
10 mechanical comprehension tests were  
11 consistent and substantial predictors of  
12 firefighter job performance.

13 Now, in the Barrett study, the  
14 authors talk about cognitive tests and  
15 mechanical tests and composites of  
16 cognitive and mechanical. What I just  
17 read from, mechanical and comprehension,  
18 do you mean composites of cognitive and  
19 mechanical tests?

20 A. I was referring, I can't speak  
21 for Mark Schemmer, but when we wrote this  
22 sentence, I am comfortable with that  
23 statement as applying to either one or  
24 both of the, either cognitive or  
25 mechanical or the composite of cognitive

1 P. Bobko

2 best I can remember, someone on that  
3 project was developing situational  
4 judgment test items to predict second tour  
5 performance.

6 Q. When was this that someone was  
7 developing situational judgment items?

8 A. I don't know exactly.

9 Q. Can you give me an  
10 approximation?

11 A. The project took place between  
12 1981 and 1991, so it would have been  
13 sometime in that period.

14 Q. And do you recall who was  
15 developing situational judgment items?

16 A. The person I think who was in  
17 charge maybe of that project was a  
18 gentleman named Wally Borman, B-O-R-M-A-N,  
19 I think, that's a guess.

20 Q. Were you involved at all in the  
21 development of the situational judgment  
22 items?

23 A. No.

24 Q. Was any validation study done on  
25 the situational judgment device that was

1 P. Bobko

2 any tests that didn't have cognitive  
3 components in them.

4 Q. Do you remember recommending any  
5 tests that weren't purely cognitive?

6 A. I can't answer your question  
7 because I don't know what purely cognitive  
8 means.

9 Q. Okay. Did you recommend the  
10 development of any selection devices that  
11 measured abilities, knowledge, skills,  
12 other than cognitive?

13 A. I don't remember recommending  
14 any such.

15 Q. Do you know whether any such  
16 devices were developed for the air traffic  
17 controller's position or any of them?

18 A. No, I don't.

19 Q. How many written selection  
20 devices have you developed for any  
21 position?

22 A. I would probably say zero. I've  
23 had a hand in the strategy of a variety,  
24 but actually developing them, none.

25 Q. How many content validity

1 P. Bobko

2 studies have you been involved in  
3 conducting?

4 A. Other than watching 6019 get  
5 developed, none. Sorry, I was just  
6 tracing through my memory.

7 Q. That's okay. I know it's been a  
8 pretty long career. How many articles  
9 have you published that discussed the  
10 requirements or procedures for content  
11 validation?

12 A. I can't answer that question. I  
13 think a lot of my work has to do with  
14 validation strategies in general, some may  
15 touch on them and some may not. I really  
16 couldn't estimate.

17 Q. Is content validation the main  
18 focus of your work?

19 A. No.

20 Q. Have you published any books  
21 that involve discussions of the  
22 requirements or procedures for content  
23 validation?

24 A. No.

25 Q. How many construct validity

1 P. Bobko

2 Q. How about job analysis?

3 A. Same answer, a little less so  
4 than adverse effect.

5 Q. Do you consider yourself an  
6 expert in employment test development?

7 A. Somewhat, somewhat is my  
8 shortcut for saying the same thing that I  
9 said for job analysis, if you don't mind.

10 Q. Do you consider yourself an  
11 expert in firefighter job tasks or duties?

12 A. No.

13 Q. Do you consider yourself an  
14 expert in firefighter or public safety  
15 selection devices?

16 A. No.

17 Q. Do you consider yourself an  
18 expert in content validation?

19 A. Again, to some extent in the  
20 whole arena of validation and how  
21 different methods might relate, but  
22 content validity per se, no.

23 Q. Do you consider yourself an  
24 expert in criterion-related validation?

25 A. Yes, or I think other people

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF NEW YORK  
Civil Action No. 07-CV-2067

-----X  
THE UNITED STATES OF AMERICA,

Plaintiff,

and  
VULCAN SOCIETY, INC., for itself and on  
behalf of its members, CANDIDO NUNEZ,  
MARCUS HAYWOOD and on behalf of a Class  
of All Others Similarly Situated,  
Plaintiff-Intervenors,

- against -

CITY OF NEW YORK, FIRE DEPARTMENT OF THE  
CITY OF NEW YORK, NEW YORK CITY  
DEPARTMENT OF CITYWIDE ADMINISTRATIVE  
SERVICES, MAYOR MICHAEL BLOOMBERG and  
NEW YORK CITY FIRE COMMISSIONER NICHOLAS  
SCOPETTA, in their Individual and  
Official capacities,

Defendants.  
-----X

271 Cadman Plaza East  
Brooklyn, New York

July 2, 2008  
9:39 a.m.

CONTINUED DEPOSITION of  
Non-Party Witness, PHILIP BOBKO, pursuant  
to Federal Rules of Civil Procedure, held  
at the above place, date and time, before  
Alice Schulman, a Notary Public of the  
State of New York.

VERITEXT REPORTING COMPANY

212-267-6868

516-608-2400

439eb014-9232-483d-a564-c3b0dad61d2b

1 P. Bobko

2 MR. FRAENKEL: She's asking you  
3 to compare, let's say the first one  
4 with the remaining six, and the second  
5 one with the remaining six, how they  
6 rate compared to each other.

7 MS. SEELEY: Exactly, thank you,  
8 Mr. Fraenkel.

9 A. So if I could rephrase for my  
10 own understanding, you're asking me within  
11 this list of seven tasks, do some rely or  
12 are some more invoking of cognitive  
13 processes than others within this list of  
14 seven, is that your question?

15 Q. That's it.

16 A. Thank you. I can't find a way  
17 to indicate that, any comparative level of  
18 cognition across these seven tasks.

19 Q. On page 28 of your report you  
20 refer to the DCAS report. When you refer  
21 to the DCAS report, you mean what we call  
22 the Morrongiello report, Exhibit 54, is  
23 that correct, Exhibit 54 should be in  
24 the --

25 A. These are your definitions.

1 P. Bobko

2 Q. I can show you a copy of it.

3 A. Yes.

4 Q. And you say that DCAS took the  
5 steps that you list in your report to  
6 develop the 1999 written exam. And just  
7 to be clear, when you refer to the 1999  
8 written exam, that's written exam 7029,  
9 correct?

10 A. That's my understanding, yes.

11 Q. In the part of your report that  
12 discusses job relatedness and business  
13 necessity, you don't talk about the later  
14 exam, written exam 2043, correct?

15 A. Correct.

16 Q. Why didn't you talk about  
17 written exam 2043 in the job relatedness  
18 and business necessity section of your  
19 report?

20 A. It was represented to me that,  
21 and I believe Mark Schemmer, that the  
22 exams were very similar in their content,  
23 very similar for their content validity  
24 and how they were constructed.

25 My review of, I remember my



1 P. Bobko

2 review of those exams seemed to confirm  
3 that they were very similar. So the  
4 report just focussed on one exam because  
5 of the assumption of similarity.

6 Q. Who represented to you that the  
7 two written exams were very similar?

8 A. The best I can remember, it  
9 would have been Ms. Pestana.

10 Q. Are written exam 7029 and  
11 written exam 2043 alternate forms?

12 A. I don't know.

13 Q. Are they parallel forms?

14 A. I don't know.

15 Q. Are they equivalent forms?

16 A. I don't know.

17 Q. In your opinion, does the  
18 Morrongiello report, what we marked as  
19 Exhibit 54, comply with the requirements  
20 in the SIOP principals for the  
21 documentation of validity information for  
22 use of an employment test?

23 A. I couldn't specifically answer  
24 your question in the sense that I don't  
25 and can't recall every line and Section 15

1 P. Bobko

2 seem to be fairly standard.

3 Q. Would you say that they used the  
4 best practices in the field in all the  
5 details of what they did?

6 A. I can't answer that question.

7 Q. Why can't you answer that  
8 question?

9 A. Because I don't know what best  
10 practice is. I don't believe the field  
11 knows what best practice is.

12 Q. Still on page 28, it's up there  
13 among the bullet points in the middle, you  
14 have a sentence that begins "The DCAS  
15 report also notes that they," and in the  
16 bullets following that you say DCAS used  
17 nine of the abilities in the Fleishman  
18 taxonomy to write items for the written  
19 exam.

20 In your opinion, of all the  
21 abilities that are required by or that are  
22 important to the entry level firefighter  
23 job, what percentage do those nine  
24 abilities represent?

25 A. I don't know.

P. Bobko

Q. Can you give me an estimate?

A. No.

Q. Why not?

A. It's very difficult to give an estimate because I don't have a recollection of all of the other abilities that were used at some point in any of the job analyses that the DCAS individuals were invoking.

Q. Well, according to Mr. Morrongiello's report, they found that 18 cognitive abilities on Fleishman's taxonomy were important to the job and they choose to measure only nine of those 18. So does that help you at all in determining how many of the important abilities the nine that they tried to measure represent?

A. I need -- I think I'm beginning to understand your question, but I need more specification if that's okay.

Q. What more do you need?

A. If you could reask the question or tell me, you're asking for a percentage

P. Bobko

Q. Have you ever given testimony to that effect?

A. Not to that specific statement that I recall.

Q. Let me ask a hypothetical. If there are ten KSAs that are important for the performance of the job, all other things being equal, if a test that measures five of them will have less content validity than a test that measures those same five plus three more; isn't that right?

A. Yes.

Q. Do you agree with the following, the content validity of an exam will be weakened if it doesn't measure as many important KSAs as possible?

A. Yes.

Q. Do you agree with the following, if you use a content validity approach to develop a test for use as a pass/fail screening device, then you would want the test to represent as many of the important KSAOs as possible?

1 P. Bobko

2 excluded from the exam, you would have to  
3 question the content validity of the exam?

4 A. No.

5 Q. Have you ever given that  
6 testimony?

7 A. Not that I can tell you.

8 Q. In the parts of your report  
9 where you talk about job relatedness and  
10 business necessity, you didn't discuss the  
11 pass/fail cutoff scores the City used on  
12 written exam 7029 and 2043, correct?

13 A. Correct.

14 Q. Do you agree with the following  
15 statement, the inferences drawn from test  
16 scores should be limited to those for  
17 which validity evidence is available?

18 A. I can't agree with that. I  
19 don't understand that sentence.

20 Q. Okay, do you agree with the  
21 following definition of validity, validity  
22 is the extent to which inference is based  
23 on test scores are justified by the  
24 evidence?

25 A. Sure.

P. Bobko

test was used in a rank ordered way when combined with the physical ability test, and therefore, and the section on business necessity speaks to the content validity of the written exam portion of that composite and, therefore, the entire section in some sense relates to the issue that you raised.

Q. As we established yesterday, that section doesn't talk at all about the PPT, does it?

A. That's correct.

Q. And you don't say anything in your report about whether the method the City used to combine the written exam scores and PPT scores was appropriate, correct?

A. Correct.

Q. Looking at page 29 of your report, the part where you talk about feasibility and business necessity, the first two points, bullet points on this page talk about the job relatedness of cognitive tests, right?

1 P. Bobko

2 A. Yes.

3 Q. There's nothing specific in  
4 those about the tests at issue here the  
5 way they were used, right?

6 A. That's correct.

7 Q. In the third bullet point, you  
8 suggested that efficiency and feasibility  
9 with respect to the costs and processing  
10 of large numbers of applicants might be  
11 one of the reasons the City used an  
12 assessment of cognitive aptitude.

13 That consideration regarding  
14 cost and processing of large numbers would  
15 also support the use of written exam 6019,  
16 wouldn't it?

17 A. Yes.

18 Q. In fact, that consideration  
19 would support the use of any objective  
20 scored written exam, not just written exam  
21 7029 or 2043, right?

22 A. No.

23 Q. How is that not right?

24 A. One could construct written  
25 exams that were allegedly objectively



1 P. Bobko

2 scored that I could imagine are quite  
3 onerous to administer and score and report  
4 the results back to applicants.

5 Q. Can you give me an example of  
6 such a test?

7 A. A written exam that takes 84  
8 days to give and has 84 times 5,000  
9 questions.

10 Q. Have you ever seen an exam like  
11 that?

12 A. No.

13 Q. Can you give me an example of an  
14 exam that you've seen?

15 A. That?

16 Q. That would not fulfill this  
17 consideration regarding cost and  
18 processing of large numbers of applicants  
19 on an objectively scored written exam.

20 A. No.

21 Q. Looking at the fourth bullet  
22 point here at the top of page 29, you  
23 suggest here that consistency in the  
24 methods and processes used to evaluate  
25 applicants might be one of the reasons the



1 P. Bobko

2 City used an assessment of cognitive  
3 aptitude. That consideration regarding  
4 consistency would equally support the use  
5 of written exam 6019, wouldn't it?

6 A. Yes.

7 Q. I would like to hand you Exhibit  
8 -- let me find it first.

9 MS. SEELEY: Why don't we take a  
10 ten-minute break.

11 (A recess was taken.)

12 Q. Dr. Bobko, I found the exhibit I  
13 was looking for. It's this one that was  
14 previously marked Exhibit 200, and for the  
15 record, it is Bates numbered DCAS 0015229  
16 through 15246, and on the front cover has  
17 the title Test Development Report for 6019  
18 Firefighter FDNY.

19 Dr. Bobko, have you seen this  
20 document before?

21 A. In its entirety, no.

22 Q. You did see drafts of the Test  
23 Development Report for 6019, didn't you?

24 A. I saw drafts of part of the Test  
25 Development Report, yes.

1 P. Bobko

2 that's been marked Exhibit 65. I'm sorry,  
3 66. Is that marked 66, what's the mark on  
4 that?

5 A. 66.

6 MR. LEVY: Are we okay on that?

7 I just misspoke.

8 Q. Do you recognize this document?

9 A. No.

10 Q. Is this a company you worked  
11 for?

12 A. It's a company whose Scientific  
13 Advisory Committee I sit on, yes.

14 Q. I'd like you to read this over  
15 and tell me if this is a summary of some  
16 of the services that this company whose  
17 advisory committee you sit on provides.

18 A. Sure, it describes some of the  
19 activities that I think that company  
20 engages in.

21 Q. Does this company know what its  
22 doing?

23 MR. FRAENKEL: Objection as to  
24 form.

25 MR. LEVY: Never mind,

1 P. Bobko

2 withdrawn.

3 Q. Here's the question. The  
4 company whose board you sit on or what's  
5 the committee you sit on, say it again,  
6 what is your exact connection again to the  
7 company?

8 A. It's called a Technical Advisory  
9 Committee, I think, something like that.

10 Q. And you are a committee member?

11 A. That's correct.

12 Q. Are the members called  
13 consultants or are they called committee  
14 members?

15 A. Committee.

16 Q. How many people sit on the  
17 committee?

18 A. About ten.

19 Q. And are they all doctorate level  
20 people?

21 A. I think all but one.

22 Q. And do they give advice to the  
23 company on testing issues?

24 A. Strategic advice, yes.

25 Q. According to this the company

1 P. Bobko

2 you have an opportunity to ask questions,  
3 to ask the teacher, to ask your  
4 colleagues, you have time to look things  
5 up generally like studying in any  
6 educational forum, but on the job you have  
7 to read something immediately?

8 A. I don't know if that distinction  
9 carries, I don't know if that distinction  
10 is accurate, I just don't know.

11 Q. Are you aware that a lot of  
12 linking panel folks testified that they  
13 understood written comprehension on the  
14 test to refer to the written comprehension  
15 you need to, in the Academy basically or  
16 to study materials, but not actually the  
17 reading level you need on the job?

18 A. No, I didn't know that.

19 Q. You didn't do any kind of a  
20 check yourself of the reading level when  
21 you determined, I'm sorry, when you wrote  
22 your report for 7029 and 2043?

23 A. That's correct.

24 Q. Do you know if, do you know what  
25 SMOG is, do you know what SMOG is

**EXHIBIT U**

**Expert Report  
of  
David P. Jones, Ph.D. & Leaetta M. Hough, Ph.D.**

**Prepared in the Matter of  
United States v. City of New York  
(Civil Action No. 07-CV-2067 E.D.N.Y)**

**Prepared Jointly  
By**

**David P. Jones, Ph.D.  
Growth Ventures, Inc.  
P. O. Box 299  
East Jordan, MI 49727  
231.536.0111  
[djones@talenttechnologycorp.com](mailto:djones@talenttechnologycorp.com)**

**and**

**Leaetta M. Hough, Ph.D.  
The Dunnette Group, Ltd.  
370 Summit Avenue  
St. Paul, MN 55102  
651.227.4888  
[leaetta@msn.com](mailto:leaetta@msn.com)**

**July 31, 2008**

importance, compared to the importance of the physical or cognitive requirements of the position.

93. In effect, the job analysis study that served as the foundation for Written Exam 7029 explored only the cognitive abilities associated with the job. This limitation in scope is not discussed in the City's job analysis report, but it severely limits the conclusions that can be drawn with respect to the requirements underlying success in the firefighter job.

94. The implications of this shortcoming are many. Collectively, they result in the City's job analysis study providing:

- a) No basis for determining the relative importance of the qualifications addressed by Written Exam 7029, as compared to the job's other physical and non-cognitive requirements;
- b) No basis for determining the relative weight to be accorded to Written Exam 7029, as compared to other components of the selection process;
- c) No basis for evaluating whether any disparate impact produced by Written Exam 7029 might be defensible based upon arguing that the requirements it measures are the job's most important requirements;
- d) No basis for establishing a cut-off or qualifying score on Written Exam 7029; and
- e) No basis for arguing that Written Exam 7029 meets the standards of content validity, because the prerequisite definition of job content was not accomplished.

95. By analogy, the City's job analysis study is akin to exploring the pilot's job by examining only those skills involved in the take-off of the aircraft; thereby recommending that a pilot selection test only involve an assessment of take-off proficiency. In such a test, requirements associated with flight planning, pre-flight inspection, navigation, systems monitoring, communications, weather tracking, safety monitoring, execution of emergency procedures and, yes, landing would go unaddressed, because the job analyst explored only those requirements associated with developing a test of take-off proficiency.

**XII. The flaws in the job analysis underlying Written Exam 7029 resulted in obtaining only limited information, all of which focused on the job's cognitive ability requirements. Many abilities or personal attributes that enable one to perform the work of a firefighter successfully were completely ignored.**

96. Other job analysis studies known to the authors, and cited below, show that the firefighter position involves a host of requirements other than cognitive abilities. All were ignored in the City's job analysis study.
97. For example, less than ten years after the job analysis study underlying Written Exam 7029, Dr. Catherine Cline's (Cline, 2007) job analysis of the New York City firefighter job revealed eight non-cognitive attributes to be highly important for performing New York City firefighter tasks. The eight (as defined in Dr. Cline's job analysis study) included:
  - **Tolerance for Stress** – the degree to which one can maintain stability of performance under pressure and/or opposition. (For example: Assist someone in cardiac arrest).
  - **Adaptability** – the ability to maintain effectiveness in varying environments, with various tasks, responsibilities, or people. (For example: Interact with other agencies such as the Police Department or Consolidated Edison. Interact with emotionally disturbed persons).
  - **Tenacity** – the ability to stay with a position or plan of action until the desired objective is achieved or is no longer reasonably attainable. (For Example: Stay at a fire scene until the fire is out, or with a medical victim until relieved).
  - **Integrity** – the degree to which one can maintain social, ethical and organizational norms in job-related activities. (For example: Enter an apartment and respect the belongings of others).
  - **Work Standards** – the ability to set high goals or standards of performance for self, subordinates, others and organization. Dissatisfied with average performance. (For example: Seek out constructive criticism. A new Firefighter asks questions on what to do next).
  - **Resilience** – the degree to which one can handle disappointment and/or rejection while maintaining effectiveness. (For example: Accept criticism and learn from mistakes. Even when faced with fatalities, learn to keep going and doing what still needs to be done).
  - **Coordination** – the ability to adjust actions in relation to others' actions. (For example: Inside team forcing door has to coordinate actions with one holding the Halligan tool and the other hitting it with an axe. If one person on a team is occupied with a rescue, another person on the team may have



to take over that person's responsibilities as well as his or her own. In administering CPR, one person may provide chest compressions, while the other person supplies oxygen).

- **Establishing and Maintaining Interpersonal Relationships** – the degree to which one can develop constructive and cooperative working relationships with others, and maintain them over time. (For example: Learn to receive constructive criticism. Your life depends on the other members of your unit performing their jobs properly; you have to learn to trust them. Shared communal activities, such as cooking and eating, help to establish and maintain good interpersonal relationships in the fire house).
98. In Dr. Cline's study, all of these attributes were rated 4.0 or higher on an "importance scale" where 3 = *Important*, 4 = *Very Important*, and 5 = *Critical* to the performance of the firefighter job. Yet, none of these personal attributes were explored in the job analysis, let alone the test design plan, the City used to support Written Exam 7029.
99. According to Dr. Cline, she specifically sought to determine whether the FDNY firefighter job had changed and concluded that it had not changed between the 1998 and 2007 job analyses. (Cline Deposition, p. 73 – 74).
100. Shown below in Exhibit 1 are the firefighter task clusters that were identified in the job analysis underlying Written Exam 7029 and that of Dr. Cline:

**Exhibit 1**

**Task Clusters Composing the Job Analysis Questionnaires  
Used in Developing Written Exams 7029 and 6019**

<b>Firefighter Job Analysis Questionnaire Task Clusters</b>	
<b>1998 Morrongiello Job Analysis (Exam 7029)</b>	<b>2007 Cline Job Analysis (Exam 6019)</b>
Initial Response to Incidents/Driving	Initial Response to Incident
Size-up	Size-up and Initial Actions
Climbing and Portable Ladder Activities	Climbing and Ladder Activities
Building Entry	Building Entry
Search	Search
Rescue	Rescue
Ventilation	Ventilation
Supplies Water for Hose Operation	Preliminary Actions Needed for Hose Operation
Hose Operations During Extinguishment	Hose Operations During Extinguishment
Overhaul	Overhaul
Salvage	
Clean Up/Pick Up	
Equipment Maintenance	Station and Equipment Maintenance/Chores
Station Duties and Chores	
Inspection of Buildings/Hydrants	Inspection of Buildings/Hydrants/Alarm Boxes
Extrication	Extrication
Providing Medical Assistance	Providing Medical Assistance
Elevator-Related Tasks	
Training	Training
Watch Duties	Watch Duties
Miscellaneous	

101. Inspection of the task clusters for the two job analyses reveals that the overall makeup of the firefighter job was essentially unchanged. In fact, the four task areas identified in the job analysis for Examination No. 7029, but not in that for Examination No. 6019, were task clusters with so few defining tasks that Dr. Cline assigned their individual tasks to other, equally sensible task clusters, thereby eliminating these categories. Thus, the absence of the task clusters Salvage, Clean Up/Pick Up, Elevator-Related Tasks, and Miscellaneous in Dr. Cline's results did not reflect changes in the job, but only a more parsimonious presentation of findings.
102. As a result, the 2007 job analysis revealed a job made up of tasks essentially similar to those on which Written Exam 7029 was based. Yet, the Cline job analysis identified eight additional attributes important for effective performance. All of these additional attributes were non-cognitive in nature. All were missing from the 1998 job

analysis results, because the City never explored anything beyond the cognitive abilities that had been analyzed in developing its most recent predecessor examination, that produced by Landy et al. in 1992.

103. In summary, the 1998 job analysis that formed the foundation for Written Exam 7029 was fundamentally flawed by its extremely narrow coverage of the job requirements used as a basis for developing a content valid written examination. Thus, the written examination produced from the analysis also was fundamentally flawed.

**XIII.** Job analyses of firefighter positions in other cities, both large and small, offer compelling evidence that attributes other than cognitive abilities are important to successful performance of firefighter work. As early as the 1970s, job analyses of firefighter jobs indicated that such attributes are important to the work of firefighters. As discussed above, these areas were not even investigated in the City's job analysis for Written Exam 7029.

104. For example, a chapter in the *Job Analysis Handbook for Business, Industry, and Government* published in 1988 described a nationwide job analysis of firefighter jobs. The project, performed in 1975 and 1976, concluded that non-cognitive attributes are important for successful performance of firefighters. Eleven of the 20 "required firefighter abilities and characteristics" reported in this nationwide study were non-cognitive in nature, as compared to only four that were cognitive, and five that were physical.
105. This work was performed "...under a contract from the U.S. Office of Personnel Management (OPM) with two functions. First, OPM wanted the study to serve as a model procedure, illustrating professional job analysis methods for personnel selection purposes, satisfying all legal requirements of the employee selection guidelines then in effect (1976). Second, it was to provide job information that local fire departments could use in developing their own firefighter selection procedures." (Bownas, 1988, p. 1255).
106. The non-cognitive attributes identified by this nationwide study were: Courage, Resistance to Stress, Teamwork, Activity, Responsibility, Desire to Learn, Getting Along with People, Honesty, Cleanliness, Medical Interests, and Construction Trade Interest. Though differing in labels, many of these are the same as those identified by Dr. Cline in her 2007 job analysis for the City of New York, as illustrated, for example, in Exhibit 2.

**Exhibit 2**

**Non-Cognitive Attributes Identified in Both a 1975-76  
Nationwide Study and 2007 City of New York Job Analysis**

<b>1975-76 Nationwide Job Analysis</b>	<b>2007 Cline New York City Job Analysis</b>
Resistance to Stress	Tolerance for Stress
Honesty	Integrity
Responsibility	Work Standards
Getting Along with People, Teamwork	Establishing & Maintaining Interpersonal Relationships

107. The Department of Labor's O\*NET (Occupational Information Network) provides another important source of information about work in hundreds of jobs and occupations in the U.S. workforce, including firefighters. O\*NET, developed in the early and mid-1990s, was developed as the replacement for the Department of Labor's Dictionary of Occupational Titles (DOT) (Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1995, 1996, 1999; Peterson et al., 2001). O\*NET provides comprehensive job descriptive information including information about knowledge, skills, abilities, and work styles. The on-line summary report for municipal firefighters found in O\*Net includes the non-cognitive skills, abilities, and work styles shown in Exhibit 3 (<http://online.onetcenter.org/link/summary/33-2011.01>).

**Exhibit 3**

**Examples of O\*NET On-line Summary of  
Non-cognitive Skills, Abilities, and Work Styles for Fire Fighters**

<b>Dependability</b>	Being reliable, responsible, and dependable, and fulfilling obligations.
<b>Cooperation</b>	Being pleasant with others on the job and displaying a good-natured, cooperative attitude.
<b>Coordination</b>	Adjusting actions in relation to others' actions.
<b>Concern for Others</b>	Being sensitive to others' needs and feelings and being understanding and helpful on the job.
<b>Service Orientation</b>	Actively looking for ways to help people.
<b>Social Orientation</b>	Preferring to work with others rather than alone, and being personally connected with others on the job.
<b>Initiative</b>	Willingness to take on responsibilities and challenges.
<b>Persistence</b>	Persistence in the face of obstacles.
<b>Attention to Detail</b>	Being careful about detail and thorough in completing work tasks.
<b>Self Control</b>	Maintaining composure, keeping emotions in check, controlling anger, and avoiding aggressive behavior, even in very difficult situations.
<b>Stress Tolerance</b>	Accepting criticism and dealing calmly and effectively with high stress situations.
<b>Active Listening</b>	Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.

108. These significant, nationwide, government-funded projects provide a long established history of the importance of several non-cognitive characteristics needed for effective performance of firefighter tasks. They illustrate the results to be expected in any comprehensive job analysis of the firefighter position.
109. Even the City's expert, Dr. Schemmer, acknowledged in his deposition testimony the importance of several of these characteristics for firefighters, and the possibility that the City could have administered a paper-and-pencil personality measure, available at the time and with small mean score differences between blacks and whites, to measure

such characteristics, increment validity, and reduce adverse impact (Schemmer Deposition, p. 280-306).

110. Most, if not all, of this job analytic information was available prior to the job analysis underlying the development of Written Exam 7029. Yet, the City's job analysis failed to explore either the physical or non-cognitive requirements of the firefighter job, a fatal flaw corrected at least to some extent in the more recent Cline (2007) job analysis.
111. Appendix D shows examples of firefighter examinations resulting from more comprehensive studies; examinations that incorporate non-cognitive elements. Most of the examinations were developed prior to 1999. The citations shown demonstrate the feasibility of measuring the broad range of job requirements associated with the firefighter job, not just those associated with cognitive demands.

**XIV. The manner in which the job analysis underlying Written Exam 7029 was executed produced several additional flaws, all limiting the range of conclusions that can be drawn from the study.**

112. Using a series of numerical decision rules, those tasks identified by the City's Examination No. 7029 job analysis sample as composing the most important duties of the job, and those cognitive abilities rated as the cognitive abilities most important to effective performance were identified. Eighteen cognitive abilities and 111 tasks met the criteria set.
113. Here, though, information required to actually interpret the job analysis results is not provided. The study failed to answer any of the following questions:
  - a) Were there any systematic differences in firefighter job duties or cognitive ability requirements as a function of duty assignment, location, participant tenure on the job, etc.? If so, what were they?
  - b) To what extent did the nearly 200 firefighters agree on the profile of job duties and cognitive ability requirements they produced in completing the survey? Were the resulting data internally consistent and statistically reliable?
  - c) When reviewed from a statistical analysis point of view, did the assignment of job tasks to task clusters "hold together," or should some of the tasks have been assigned to clusters with which they shared more commonality in importance or frequency of performance?



- d) Were any of the 111 tasks and 18 cognitive abilities identified as important to the firefighter job universally important, or did certain groups of firefighters (locations, assignments, experience levels, etc.) rate them as critically important, whereas other groups of firefighters rated them as less important, resulting in average importance ratings sufficiently high so as to misleadingly label the tasks or abilities as "generally important?"
- e) Were all of the cognitive abilities identified as important to the job actually important on "Day 1," that is, were all of the abilities those that a new recruit must be capable of demonstrating in a fully effective fashion on their first day of employment, as opposed to acquiring them through training, experience, and practice?

114. The data necessary to answer many of these questions were collected and available for review. As executed, though, the job analysis study offered none of these reviews. Nor did the City's experts offer any such reviews. Analyses as fundamental as those needed to be confident in the basic statistical reliability of the job analysis data were not reported.

**XV. "Linking judgments" provided by the City's panel of 12 firefighters were used, along with information obtained from the job analysis survey, to determine how many exam questions should be included in Written Exam 7029 for each of the nine cognitive abilities. Troubling in this stage of the project is information regarding the degree to which the 12 firefighters involved in producing the final written examination specification did not understand and perform their assignment.**

- 115. After administering the job analysis questionnaire, City staff asked a group of 12 firefighters to review the list of 111 "important tasks," grouped into 21 task categories (or "clusters"), and to judge the extent to which each of the 18 "important abilities" was required to perform each of the 21 "important task clusters."
- 116. At a minimum, this activity called for each of the 12 participating firefighters to make 378 judgments regarding the importance of each task category-to-ability relationship. Maximally, it required them to consider nearly two thousand (111 multiplied by 18) combinations of important tasks and important abilities. In either case, the 12 firefighters were asked to perform a task that would be highly demanding, even for individuals trained in job analysis, test design, psychological measurement, or human performance assessment. The group of 12 firefighters possessed no such training.

117. For example, the 12 firefighters were asked individually to judge the degree to which each of the 21 important task clusters required:

*Flexibility of Closure: The ability to identify or detect a known pattern (like a figure, word or object) that is hidden in other material.*

118. This activity requires a noteworthy degree of knowledge regarding the underlying determinants of human performance, knowledge unlikely to fall within the training or experience of a firefighter or one who supervises firefighters. As a general rule, individuals asked to provide such judgments are trained in industrial psychology, tests and measurements, developmental psychology, or related fields.
119. In asking the linking panel to make these judgments, the City also counted on their results to accomplish a step critical to arguing the content validity of Written Exam 7029 – it used the linking panel judgments (along with ratings from the job analysis survey) to determine the number of questions that should be written for each of the nine cognitive abilities.
120. While this step might not have been a central job-relatedness issue if the City had undertaken a criterion-related validation of Written Exam 7029, it was critical if any argument of content validity is to be made. As executed, it is a key link in any argument that the content of the examination parallels the content of the job in a meaningful way.
121. There is ample evidence that those performing the linking analysis did not all understand, or follow, their assignment. First, for example, five of the 12 firefighters who were linking panel members linked virtually every job task cluster with every one of the 18 cognitive abilities. Although some of the five used the lowest, “1 – Somewhat Important” judgment in making these linkages, it strains reason, for example, that:
- *Climbing and Portable Ladder Tasks* require *Written Expression Ability*,
  - *Clean Up/Pick Up Tasks* require *Oral Expression Ability*,
  - *Watch Duty Tasks* require *Spatial Orientation Ability*, or
  - *Salvage Tasks* require *Fluency of Ideas*.
122. In addition, review of the deposition transcripts for linking panel members provides further information concerning their lack of understanding of the task with which they were presented. For example, the deposition transcript of the individual coded as Rater #1, one of the linking panel members indicates that this individual:



- Did not understand the definition of the ability Problem Sensitivity (Rater #1 Deposition, p. 57).
- Did not understand the definition of the ability Inductive Reasoning (Rater #1 Deposition, p. 58).
- Did not understand the definition of the ability Visualization (Rater #1 Deposition, p. 61-62).
- Did not understand the definition of the ability Time Sharing (Rater #1 Deposition, p. 62).
- Disagreed with her own linkage of the ability Written Comprehension for seven of the eight task clusters on which she was questioned during deposition. "Because, again, it has nothing to do with the written word to perform these tasks." (Rater #1 Deposition, p. 69).

123. During deposition, the linking panel member coded as Rater #3:

- Regarding his ratings of the importance of the ability Inductive Reasoning, stated, "I probably didn't know what it meant, so I gave it a two since I didn't know what it was, quite frankly." (Rater #3 Deposition, p. 67).
- Regarding his ratings of the importance of the ability Deductive Reasoning, stated, "I don't believe I really knew what deductive reasoning was at the time of the examination." (Rater #3 Deposition, p. 72).
- Regarding his rating of the link between the ability Memorization and the task cluster Salvage, stated, "I can't give you an honest answer as to why I gave it a three." (Rater #3 Deposition, p. 81 - 82).
- Regarding his rating of the link between the ability Written Expression and the task cluster Providing Medical Assistance, stated, "So written expression I guess would be, there's no written expression. I guess it would be irrelevant to the situation of medical...I was looking at the wrong one. I would change the answer to zero." (Rater #3 Deposition, p. 91 - 92).
- Regarding his rating of the ability Deductive Reasoning to the task cluster Training, stated, "I didn't have an explanation of what deductive reasoning was, so to answer it I was just guessing." (Rater #3 Deposition, p. 93 - 95).

124. During deposition, the linking panel member coded as Rater #10:

- In discussing the importance of the ability Inductive Reasoning, responded, "I'm not sure. It's almost the opposite of deductive, right? Can somebody explain that to me, inductive, deductive?" (Rater #10 Deposition, p. 47 – 49).
- Again, "I really wasn't sure what they're trying to say by inductive reasoning to begin with." (Rater #10 Deposition, p. 49).
- In discussing the importance of the ability Spatial Orientation, responded, "This sounds like a – it's more of a physical thing, like I mean is it saying that is it important that firefighters wear their glasses? Is that what it's saying?" (Rater #10 Deposition, p. 51 – 52).
- Regarding his rating of the linkage between the ability Visualization and the task cluster Watch Duties, stated, "It's really not that important. Like I said, I put somewhat because I don't like zero." (Rater #10 Deposition, p. 94 – 95).

125. These, and many other comments gleaned from the linking panel members' deposition transcripts, suggest that the panel members were not equipped or assisted in gaining the understanding needed to perform their assignment.

126. Notably, the linking panel with which Dr. Cline worked in 2007, appears to have been conducted with a greater degree of oversight, explanation of the task to the panel members, quality control, and structure.

127. In summary, though linking panel members appear to have experienced problems in performing the task they were presented, their judgments were used to determine the number of questions that would be used to assess each of the cognitive abilities measured by the new written examination. In our professional opinion, this represents a fatal flaw in the information used to determine the test development plan for Written Exam 7029.

128. Further, if the linking panel members did not understand the definitions and meanings of the cognitive abilities with which they worked in executing their assignments, another question is raised: Did the sample of firefighters who completed the initial job analysis questionnaire have the same difficulty in understanding the cognitive abilities they rated with respect to importance in performing the firefighter job?

129. The answer to this question is unknowable, but the linking panel deposition transcripts raise serious questions in this regard. If the answer is "no," then the veracity of the overall job analysis data base on which Written Exam 7029 was based comes into question.

**XVI.** The approach followed in developing Written Exam 7029 produced an examination that bore little relation to the range of abilities a) required by the job and b) amenable to paper-and-pencil testing. Not only did it fail to address any of the job's non-cognitive requirements, it failed to measure important cognitive abilities that could have been addressed by a paper-and-pencil examination.

130. Using the results of the job analysis, the City determined that only nine of the 18 abilities identified as important to the job could be assessed by Written Exam 7029, leaving the other one-half to go unaddressed.

131. The result of this decision was to reduce further the overlap between Written Exam 7029 and the overall range of requirements important to success as an FDNY firefighter – even when considering only its cognitive ability requirements.

132. Those cognitive requirements retained for consideration in designing the exam included only: *Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation, and Visualization.*

**XVII.** In developing the City's 2007 written examination, Dr. Cline undertook development of a means to measure 20 of the cognitive and personal attribute requirements her job analysis had identified as important to performance of the FDNY firefighter job; 18 directly, and two (Written Comprehension and Time Sharing) indirectly. Abilities cited by the City as not amenable to paper-and-pencil measurement when it designed Written Exam 7029 were addressed in the Cline exam.

133. Dr. Cline (2007) determined that, of the nine cognitive abilities eliminated from the design of Written Exam 7029, three could be measured directly and one could be measured indirectly through a written test. One of the nine cognitive abilities (Written Expression) the City set out to measure in Written Exam 7029 was determined by Cline as not sufficiently important to the job to be included in the new examination.

134. This, of course, does not take into account the additional non-cognitive abilities that Dr. Cline also added to her test development plan in order to better cover the broad range of the job's requirements.

135. Exhibit 4 below contrasts the test design decisions made by the City for Written Exam 7029 with those of Dr. Cline in creating the new Written Exam 6019. The left column of the exhibit shows the nine cognitive abilities addressed by Written Exam 7029, as well as the nine abilities (shown in italics) excluded from its exam development efforts. The right column shows the cognitive abilities included by Dr. Cline in developing the new Written Exam 6019.

**Exhibit 4**

**The Examination Development Plans Associated with Written Exams 7029 and 6019**

<b>Morrongiello Test Development Plan (1999)</b>	<b>Cline Test Development Plan (2007)</b>
<b>Cognitive Abilities Selected for Written Testing</b>	<b>Cognitive Abilities Selected for Written Testing</b>
Written Comprehension	Written Comprehension (Indirectly)
Written Expression	
Memorization	Memorization
Problem Sensitivity	Problem Sensitivity
Deductive Reasoning	Deductive Reasoning
Inductive Reasoning	Inductive Reasoning
Information Ordering	Information Ordering
Spatial Orientation	Spatial Orientation
Visualization	Visualization
	Number Facility
<b><i>Not Selected for Written Testing</i></b>	
<i>Oral Comprehension</i>	
<i>Oral Expression</i>	
<i>Fluency of Ideas</i>	
<i>Originality</i>	
<i>Speed of Closure</i>	Speed of Closure
<i>Flexibility of Closure</i>	Flexibility of Closure
<i>Perceptual Speed</i>	Perceptual Speed
<i>Selective Attention</i>	
<i>Time Sharing</i>	Time Sharing (Indirectly)

136. Counting the number of cognitive abilities addressed in common by both examinations (8), versus the number of cognitive abilities unique to one examination or the other (6), yields an index of the two examinations' overlap; that is, there is only approximately 57 percent ( $8 / 8 + 6 = .57$ ) overlap in the cognitive abilities the two test developers set out to measure.

137. Adding to these values the seven personal abilities that Dr. Cline included in her test design plan – *Adaptability, Tenacity, Integrity, Work Standards, Resilience, Coordination, and Establishing and Maintaining Interpersonal Relationships* – results in an overlap between the two written exams of only about 38 percent ( $8 / 8 + 13 = .38$ ) in the abilities the two test developers set out to measure.
138. The narrowness in covering the firefighter job's most important abilities inherent in Written Exam 7029 is underscored even further in Exhibit 5 below. Here, the cognitive and personal abilities deemed sufficiently important to be included in the linking panel processes for Written Exams 7029 and 6019 are presented.
139. The list of abilities is rank-ordered according to the average importance rating awarded in the Cline (2007) job analysis sample. Where the average importance ratings are the same to the first decimal place, they are listed alphabetically within that value. Abilities targeted in designing Written Exam 7029 are noted, as are those targeted in developing Written Exam 6019. The two areas cited by Cline (2007) as being measured indirectly by Written Exam 6019 are so labeled, as well.
140. Of note, the abilities identified by Cline (2007) as the focus for examination development efforts differed little from one another in the average ratings they received from the job analysis survey sample on the five-point "importance" scale. Average values ranged from the highest at 4.37 for Coordination, to the lowest at 4.00 for Resilience; virtually equal in importance.

**Exhibit 5****The Abilities Addressed in Developing Written Exam 7029 and 6019**

<b>Ability</b>	<b>Type of Ability</b>	<b>Targeted In Developing Exam 7029</b>	<b>Targeted In Developing Exam 6019</b>
Oral Comprehension	Cognitive		
Coordination	Personal		Yes
Oral Expression	Cognitive		
Spatial Orientation	Cognitive	Yes	Yes
Tenacity	Personal		Yes
Tolerance for Stress	Personal		
Written Comprehension	Cognitive	Yes	Yes (Indirect)
Deductive Reasoning	Cognitive	Yes	Yes
Inductive Reasoning	Cognitive	Yes	Yes
Information Ordering	Cognitive	Yes	Yes
Integrity	Personal		Yes
Interpersonal Relationships	Personal		Yes
Problem Sensitivity	Cognitive	Yes	Yes
Speed of Closure	Cognitive		Yes
Time Sharing	Cognitive		Yes (Indirect)
Flexibility of Closure	Cognitive		Yes
Memorization	Cognitive	Yes	Yes
Perceptual Speed	Cognitive		Yes
Visualization	Cognitive	Yes	Yes
Work Standards	Personal		Yes
Adaptability	Personal		Yes
Fluency of Ideas	Cognitive		
Number Facility	Cognitive	n/a	Yes
Resilience	Personal		Yes
Written Expression	Cognitive	Yes	

141. The exhibit underscores the relative importance of the non-cognitive areas addressed by Written Exam 6019. These attributes, part of the "content domain" of the firefighter job, go unaddressed in Written Exam 7029, as do cognitive abilities that were rated equivalent in importance to those addressed by the examination.
142. Demonstrably, Dr. Cline's experience also indicates that the City could have addressed (but did not) a broader range of the job's cognitive and non-cognitive ability requirements. The result: Written Exam 7029 offered content that overlapped only to a relatively minor degree with the full range of cognitive abilities required by the job; and even less with the combined cognitive and non-cognitive abilities addressed by Dr. Cline in Written Exam 6019.



143. In attempting to support Written Exams 7029 and 2043, the City's experts, Drs. Bobko and Schemmer, opine that:

"It . . . also is clear that many cognitive demands are placed upon a firefighter. For example, mis-evaluation of priorities at a fire scene, or inappropriately carrying out fire suppression and extinguishing activities, can lead to extraordinary costs..." (Bobko & Schemmer, 2008, p. 27).

144. They go on to opine that:

"The content of [certain] firefighter job tasks (e.g., use of information to determine, communicate, evaluate, and select things) clearly invoke cognitive processes." (Bobko & Schemmer, 2008, p. 28).

145. We agree with Drs. Bobko and Schemmer that cognitive abilities play a role in performing the firefighter job. However, Drs. Bobko and Schemmer admit they did no systematic analysis to determine whether Written Exams 7029 and 2043 measure the nine specific abilities the City set out to measure. In fact, in his deposition, Dr. Bobko indicated that the Bobko and Schemmer (2008) report was not addressing the nine cognitive abilities the written examinations set out to measure (Bobko Deposition, p. 293-294) when the authors concluded:

"In sum, the cognitively-based written exams were developed following standard job analytic and test development procedures – thus speaking to their job relatedness." (Bobko & Schemmer, 2008, p. 28).

146. The admonition of testing expert Dr. Robert Guion (1998) bears repetition in the context of evaluating the opinions of Drs. Bobko and Schemmer, and in reaching conclusions regarding the content validity of Written Exam 7029. As Guion indicates:

"Part of the argument of job-relatedness is that nothing substantial has been overlooked." (Guion, 1998, p. 61).

147. In our professional opinion, the Written Exam 7029 test development study "overlooked substantial things," including cognitive abilities both relevant and amenable to measurement in an entry-level firefighter examination, as shown by Dr. Cline in the City's most recent test development effort for Written Exam 6019.

148. Moreover, the expert report of Drs. Bobko and Schemmer (2008) is silent regarding the job-relatedness of non-cognitive factors in performing the firefighter tasks they cite. For example, the City's experts gave no consideration to the fact that non-cognitive abilities,

such as Dr. Cline's *Adaptability*, *Resilience*, and *Coordination*, play a role in performing firefighter tasks, including the tasks Drs. Bobko and Schemmer (2008) cite in their expert report.

149. In his deposition, however, Dr. Bobko acknowledged that factors other than cognitive abilities are important to the firefighter job:

Q: What is the relative importance of physical abilities, cognitive abilities and personal attributes to the successful performance of the FDNY entry level firefighter job?

A: I don't know. I can tell you they are all important based on the job analysis [done for Examination No. 6019], but that's all I can tell you. (Bobko Deposition, p. 45-46.)

150. In alluding to the feasibility of addressing the non-cognitive requirements of the firefighter job, Bobko and Schemmer state:

"Particularly given the large numbers of applicants in New York City, the cost of administering and objectively scoring other types of exams at this stage appears to have been prohibitive and not feasible (e.g., structured interviews or work sample examples would be difficult to develop, administer, score and keep secure and equivalent across all applicants)." (Drs. Bobko & Schemmer, 2008, p. 29).

151. Again, however, in developing Written Exam 6019, Dr. Cline (2007) demonstrated that easily administered, objectively scored, secure approaches to measuring the job's most important cognitive and non-cognitive requirements are available and, in fact, were available at the time Written Exams 7029 and 2043 were developed. Indeed, in his deposition, Dr. Bobko agreed that Written Exam 6019 met each of the administrative, scoring, and security objectives cited. (Bobko Deposition, p. 329 - 332.)

152. It is our understanding based upon the transcript of Dr. Cline's deposition that Dr. Bobko participated in her job analysis and examination development work for Written Examination 6019 coincident with his involvement as an expert in this matter.

153. It is our professional opinion that the job analysis and test development work underlying Written Exam 7029 fails to meet the requirements of sound, professional practice. It provides no convincing basis to conclude that the content of Written Exam 7029 rests on a thorough analysis and understanding of the variety of abilities and attributes important to performance of the FDNY firefighter job. The linking



panel judgments on which its examination development plan was based appear to have been done without sufficient understanding on the part of the linking panel members. The job analysis work reported in the Morrongiello report offers insufficient evidence to establish content validity.

154. We have reviewed the plaintiff United States' expert report of Dr. Irwin Goldstein, also concluding that there is insufficient evidence of content validity to support the City's procedures, and we agree with the conclusions reached by Dr. Goldstein.

**XVIII. The "job analysis" work underlying development of Written Exam 2043 was executed even less effectively than that for Written Exam 7029. The minimal work performed here produced a self-fulfilling conclusion that earlier job analysis findings should be taken at face value in designing Written Exam 2043. It provides no basis for reaching a conclusion of content validity for the resulting written examination.**

155. The 2002 project undertaken to develop Written Exam 2043 did not involve a job analysis study, even one as narrowly focused as that conducted by the City in 1997-1998 to develop Written Exam 7029.

156. Instead, the project undertaken by the City to develop Written Exam 2043 can be characterized best as a "test item writing project." It offers no evidence of content validity.

157. As a result, the examination suffered the same shortcomings as Written Exam 7029 with respect to addressing only nine cognitive ability areas. Thus, Written Exam 2043 also fails the same comparisons reported above with respect to the most recent written examination, Written Exam 6019, designed by Dr. Cline.

**XIX. In fact, analysis of psychometric data produced by the administration of Written Exams 7029, 2043 and 6019 demonstrates that the more comprehensive job analysis and examination design work performed by Dr. Cline resulted in an examination fundamentally different from its two predecessors.**

158. As noted, Written Exam 6019 was based upon a much broader analysis of the firefighter job than were Written Exams 7029 and 2043. Design of Written Exam 6019 also was based upon Dr. Cline's conclusions that areas considered not amenable to paper-and-pencil assessment in the earlier examinations were, indeed, so measureable.

159. Psychometric analysis of data produced through administration of the three written examinations demonstrates that the decisions made by Dr. Cline resulted in a test that not only looked different than its predecessors; it produced different results.

examinations either can be considered to measure different attributes, or some other influence is affecting the scores of candidates on one test or the other.

171. The extent to which an expectation of similar rank ordering among candidates holds can be assessed statistically by computing the correlation between scores attained by a group of candidates completing multiple examinations. The magnitude of this correlation can be used to express the strength of the relationship – the similarity in candidate rank orders – between the examinations being compared.
172. In the current matter, review of the data bases provided to us for the three firefighter examinations showed that 2,667 of the candidates who completed Written Exam 7029 also sat, approximately three years later, for Written Exam 2043. A total of 479 who sat for Written Exam 7029 also sat for Written Exam 6019 in 2007. Exhibit 7 below shows the number of candidates who sat for other possible combinations of examinations.

**Exhibit 7**

**Candidates Sitting for Written Exams  
7029, 2043, and 6019**

Examination No.	Total No. Taking Exam at Left	No. Taking Exam at Left and 2043	No. Taking Exam at Left and 6019
7029	17,145	2,667	479
2043	17,817	n/a	2,240
6019	22,056	2,240	n/a

173. According to Nunnally (1978), the correlation between two forms of a test should be no more than .20 smaller than the value of the test's internal consistency reliability coefficient in order for the two tests to be considered alternate forms of the same content. In other words, if a given test shows an internal consistency reliability of .80 and correlates with another test at a value of .65, then the two tests can be considered alternate forms under Nunnally's (1978) definition; .65 is within .20 of .80 (i.e.,  $.80 - .65 = .15$ ). If the two tests had correlated only .55 with one another, the decision would be different; .55 is not within .20 of .80.
174. In this case, the item analysis computer printouts provided by the City for the AM and PM administration sessions of Written Exams 7029 and 2043 provided information concerning the internal consistency reliability (KR-20) of the two exams. For the AM and PM administration sessions of Written Exam 7029, the candidate groups both produced internal consistency reliability (KR-20) values of .91.

187. In our professional opinion, the more comprehensive job analysis and examination design work undertaken by Dr. Cline in developing Written Exam 6019 shows the lack of job content coverage in the two predecessor examinations. This further supports a conclusion that Written Exams 7029 and 2043 do not demonstrate content validity for screening entry-level firefighters.

**D. In turning its attention to evidence of criterion-related validity, the report of the City's experts proposes that other jurisdictions' firefighter test validation studies support the use of written cognitive ability tests by the City under a "validity generalization" argument. The single validity generalization research study for firefighters that the City's experts cite actually shows contrary findings.**

**XX. In their report, defendant's experts appear to propose that "validity generalization" supports the City's use of Written Exams 7029 and 2043. They apparently argue that meta-analysis, a quantitative approach to summarizing the results of multiple criterion-related validity studies, supports a criterion-related validity argument for the two written examinations. They are incorrect.**

188. Meta-analysis is a set of methodologies that researchers use to explore and summarize results across many research studies. The method seeks to identify the overall conclusions that can be said to emerge from a cumulative body of research reflected in individual studies.

189. Validity generalization is a specific form of meta-analysis. In validity generalization, correlations between predictor (selection procedure) scores and criteria (job performance measures) assembled from multiple criterion-related validity studies are summarized in an attempt to learn whether the individual studies' patterns of results can be said to "generalize" across the studies reviewed and, potentially, to situations where no specific study has been conducted; e.g., to the City's use of a written cognitive ability test in screening firefighter candidates when no local criterion-related validity study has been completed.

190. In their expert report, Drs. Bobko and Schemmer reference a meta-analysis of entry-level firefighter validation studies prepared by Barrett, Polomsky, and McDaniel (1999). The study purports to summarize correlations (criterion-related validity coefficients) between cognitive ability tests and measures of job performance of firefighters.

191. In fact, the Barrett et al. (1999) study is the only criterion-related validity information that Dr. Bobko reported considering in forming his

conclusions about Written Exam 7029 and 2043, as indicated during his deposition:

Q: The only criterion-related evidence you talk about for firefighters is that Barrett article, right?

A: Correct. (Bobko Deposition, p. 201 - 202).

192. When questioned again:

Q: Let me make sure of one thing. In your report, you didn't cite any other criterion study on firefighters; is that right?

A: That's correct. (Bobko Deposition, p. 215).

193. Drs. Bobko and Schemmer conclude that the Barrett et al. (1999) study provides evidence of validity generalization (job-relatedness) for written cognitive ability examinations. In essence, Drs. Bobko and Schemmer argue that positive findings of criterion-related validity studies for other tests should be considered to "generalize" to the City's use of Written Exams 7029 and 2043.

194. In considering the expert report of Drs. Bobko and Schemmer, it is important to understand how meta-analysis actually works. The procedure involves computing an average criterion-related correlation coefficient, based upon the individual validity coefficients reported across all the studies a researcher, such as Barrett et al. (1999), has collected.

195. In each of the individual validity studies included in such a meta-analysis, test developers have determined whether scores attained on their tests correlate with measures of actual on-the job performance. They use correlation coefficients - sometimes referred to as "validity coefficients" - which are computed to express the direction and strength of this relationship. Correlation coefficients range between -1.00 and +1.00.

196. For example, a correlation (validity) coefficient with a value of .00 (zero) between candidates' test scores and subsequent measures of on-the-job performance indicates that there is no relationship between scores on the test and on-the-job performance. In this case, the test would be said to show no (zero) evidence of criterion-related validity. It would be said to have no value as a tool for predicting candidates' subsequent on-the-job performance.

197. A correlation (validity) coefficient greater than zero between test scores and, for example, supervisory ratings of performance effectiveness would indicate that there is a positive relationship between scores on the test and subsequent on-the-job performance.

198. Correlation (validity) coefficients also may be less than zero. A negative correlation between test scores and, for example, the number of disciplinary actions in which an employee has been involved actually reflects positively on the test. The higher the test score, the fewer disciplinary actions. Of course, observing a correlation (validity) coefficient of less than zero when a positive one is logically expected would raise questions about the criterion-related validity of a test.
199. In conducting criterion-related validity studies it is essential to determine whether the correlation (validity) coefficient for a given set of data is sufficiently different from zero to conclude that use of the test is "valid." To this end, researchers generally adopt statistical standards for determining whether a given correlation (validity) coefficient, computed for a given set of data, can be viewed as "statistically significant." These standards actually amount to determining whether the correlation (validity) coefficient may be considered "significantly different from zero."
200. Meta-analysis moves beyond this study-by-study review of validity (correlation) coefficients, and seeks to identify overall patterns that apply across a large number of studies. The meta-analysis model assumes that any variation around the average correlation computed across many individual studies is due to a) the fact that the various research study samples come from different settings with different types of measurement errors that affect the predictor and criterion measures on which the studies' validities are computed, or b) real differences (in the predictor, the job the applicants, etc.) that lead the validities to actually be different in different settings. Meta-analysis tries to explain the variance as due to sample differences and typical measurement errors.
201. Thus, the smaller the variation around the average correlation, and the larger the size of the average correlation computed in a meta-analysis study, the greater the probability that one can conclude that "validity generalizes" across all the studies summarized and, by inference, to other similar settings.
202. If the average validity is small or the variation is great and cannot be explained by typical measurement error or sample size differences, then the meta-analysis must conclude that "validity does not generalize" across the studies summarized, much less to other similar settings.

**XXI.**

As in any analysis, the confidence placed in the findings of a given meta-analysis study is a function of the comprehensiveness and objectivity with which it seeks out and assembles the individual studies on which it is based. No bias can be present in the choice of studies that are selected for the analysis, lest the study's overall conclusions be similarly biased. The Barrett, Polomsky, and McDaniel (1999) study fails in this regard.



203. Meta-analysis serves to summarize as many appropriate, independent research studies as its author can collect. While conditions (the size of the studies, their dates of execution, the metrics involved, etc.) can be set on the specific studies entered into a given meta-analysis, little pre-selection typically takes place, lest such selectivity bias the conclusions of the overall analysis. This, of course, means that studies of widely differing magnitude, quality, and impact on a given area of inquiry are treated as "equals" in such an analysis. Poorly done research is not differentiated from research done well in aggregating meta-analysis data bases.
204. It is well known, too, that research which produces findings of little note (e.g., criterion-related validation studies failing to find evidence of statistically significant validity) often is not published and, hence, does not come to the attention of researchers conducting meta-analysis studies. In the professional literature, this eventuality is referred to as "the file drawer phenomenon," (See, for example, Rosenthal, 1979; Roth, Bevier, Bobko, Switzer, and Tyler, 2001)) with the label illustrating the fact that studies with non-consequential findings often "disappear into the file drawer," rather than being published for others to find in the course of a meta-analysis review.
205. Any bias toward only those studies that show statistically significant results being published (and located by a meta-analysis researcher) would be expected to bias meta-analyses of criterion-related validity research, leading them to show greater evidence of "validity generalization" than might actually be true.
206. Hence, meta-analysis researchers are admonished to search diligently, and to include all research studies possible in their reviews. In meta-analysis, professional practice calls for locating as many individual studies as possible, whether their individual findings have been supportive of validity for a given type of selection procedure or not.
207. In this regard, even casual inspection of the Barrett et al. (1999) study reveals a number of criterion-related validity studies missing from what is represented as a "comprehensive review and meta-analysis" (Barrett et al., 1999, p. 507).
208. Without undertaking a comprehensive search for validity studies of selection tests for firefighters, the authors were easily able to locate studies predating, yet missing from, the Barrett et al. (1999) meta-analysis, including:
  - Morris & McDaniel (1989). *Report on the Validation Study for the IPMA National Firefighter Examination*. Alexandria, VA: International Personnel Management Association.

- Bownas, D. A., Rosse, R. L., & Dunnette, M. D. (1977). *Construct Validation of a Selection Battery for the Entry Level Firefighter Position*. Personnel Decisions Research Institute Report 15. Washington, DC: U.S. Civil Service Commission.
- Heckman, R. W. (1973). *Saint Paul Firefighter Validation Study*. (Report Submitted to The City of Saint Paul). Minneapolis, MN: Personnel Decisions, Inc.
- Arvey, R. (1971). *Report on Test Validation Study: Minneapolis Civil Service Firefighter Jobs*. Minneapolis, MN: Personnel Decisions, Inc.

209. Further, on review, all four of these studies included cognitive ability tests and mechanical ability tests, just as those Barrett et al. (1999) set out to analyze. Importantly, in all four studies the criterion-related validities reported were lower than the average validities reported by Barrett et al. (1999). In several instances, the validities were dramatically lower. The "comprehensive review" reported by Barrett et al. (1999) did not incorporate these findings.
210. Further, in their study summary (abstract), Barrett et al. (1999) claim their analyses were based on data from 101 samples (Barrett et al., 1999, p. 507). A review of their Appendix B (Barrett et al., 1999, p. 513) reveals that the data actually came from a total of 25 samples; correctly stated, the 25 samples produced 101 validity coefficients.
211. Drs. Bobko and Schemmer similarly report the number of samples as "more than 100 samples" (Bobko & Schemmer, 2008, p. 27), again overstating the size and scope, and thus the significance, of the Barrett et al. (1999) study.
212. In addition, Barrett et al. (1999) report: "For job performance criteria, data from 73 independent samples comprising 9,515 individuals were obtained." (Barrett et al, 1999, p. 509).
213. Again, Barrett et al.'s (1999) representation of their meta-analysis sample is misleading. For example, reference to the report's Appendix B shows that, in truth, there were no more than 25 independent samples in the entire study.
214. Dr. Bobko appeared to recognize the mischaracterization of the Barrett et al. (1999) study when a copy of the Barrett et al. (1999) publication was presented to him as Exhibit 620 during his deposition. For example:

Q: Would you look at the last page of this Exhibit 620 and see Appendix B at the bottom of that page. Do you see that?

A: Yes.

Q: How many samples does this indicate the author used?

A: 25. I count 25.

Q: So it wasn't over 100, right?

A: 25 is certainly less than 100. (Bobko Deposition, p. 204 - 205).

215. Further, nineteen of the study's 101 correlation coefficients were drawn from one sample; a criterion-related validation study executed in Columbus, OH (Kriska & Hines, 1984). Hence nearly twenty percent of the data base summarized by Barrett et al. came from a single study. Another approximately 20 percent was drawn from only two additional samples (Murdy & Norton, 1972; Psychological Services, Inc, 1983). During his deposition, Dr. Bobko acknowledged this additional shortcoming:

Q: Now, if you look back at Appendix B, for one of the studies the authors say, and I believe this is the Kriska & Hines, there were 19 r's from one sample, right, do you see that?

A: Yes, I do.

Q: So for all we know, the 24 r's listed in Table 1 for cognitive tests all came from no more than a handful of independent samples, right?

A: Again, I can't answer that question if there were independent subsamples.

Q: Does it say that anywhere in the report?

A: It doesn't say that anywhere.

Q: The issue of independent samples is pretty important for meta-analysis, right?

A: I think so. (Bobko Deposition, p. 210 - 211).

216. Barrett et al. (1999) describe the studies as having been "conducted over the past two decades" (Barrett et al., 1999, p. 509), which would include 1979 to 1999. Dr. Bobko agreed during his deposition that "the last two decades" would be considered to include 1979 to 1999. (See Bobko Deposition, p. 202 - 203).

217. Yet an examination of Barrett et al.'s Appendix B reveals that the studies were conducted between 1958 and 1986. None of the studies



was conducted in the 1990s; the majority of the studies were conducted prior to 1979. Hence, most are today over 30 years old.

XXII. The Barrett et al. (1999) meta-analysis study summarized criterion-related validities separately for cognitive tests, mechanical comprehension tests, and tests that combined both cognitive and mechanical components. The City's written examinations do not include a mechanical comprehension component. They are tests of cognitive ability, as the City's experts indicated. Contrary to the report of Drs. Bobko and Schemmer, the Barrett et al. (1999) study reveals that such tests do not show evidence of validity generalization in predicting firefighter job performance.

218. Meta-analysis is used inferentially to accept or reject the hypothesis that there is no relationship between the variables (e.g., between tests scores and measures of performance) being studied. In the Barrett et al. (1999) study, meta-analysis was used to accept, or reject, the hypothesis that cognitive ability tests are correlated with the job performance of firefighters.

219. The statistic that is computed through meta-analysis techniques to guide this decision is referred to as a "credibility value" (CV). The CV traditionally is determined by computing the "90 percent CV". The simple interpretation of the 90 percent CV is that, based upon all the studies summarized in the meta-analysis, one can conclude with 90 percent certainty that the "true validity" of the selection procedure the researcher is investigating (here, cognitive examinations) falls at or above the 90 percent CV value.

220. Since the task of a meta-analysis is to determine whether the true validity is greater than zero, researchers frequently refer to "90 percent credibility" that the true validity is greater than zero.

221. Barrett et al. (1999) found the 90 percent CV for the validity of cognitive ability tests in predicting the job performance of firefighters was -.03. Thus, the estimated population distribution of validity coefficients was so large ( $\sigma_p = .35$ ) that the 90 percent CV was below zero.

222. During his deposition, Dr. Bobko acknowledged this finding:

Q: Based on that negative -.03, wouldn't it be fair to say one cannot have 90 percent confidence that the true validity is different from zero?

A: Sure, for cognitive ability tests. (Bobko Deposition, p. 214).

223. In short, the validity of cognitive ability tests for predicting firefighter job performance, as reflected in the Barrett et al. collection of studies, does not generalize – it could not be stated with 90 percent confidence that the true validity was greater than zero.
224. Although, in his deposition Dr. Bobko attempted to argue that Written Exams 7029 and 2043 could possibly be characterized as measuring a combination of cognitive and mechanical abilities, rather than as cognitive ability tests, his assertion was based on an apparent misreading of one sentence in the Barrett et al. (1999) report stating that combinations of spatial and mechanical tests were considered mechanical for purposes of the meta-analysis. (Bobko Deposition, p. 215 – 219).
225. Dr. Schemmer, who had drafted the section of the City's expert report addressing the Barrett et al. (1999) study, correctly admitted that Written Exams 7029 and 2043 do not include mechanical ability items. (See Bobko Deposition, p. 215 – 219 and Schemmer Deposition, p. 236 – 237). They are cognitive tests.
226. Moreover, even use of the 90 percent CV may result in unacceptably high error rates; in other words, it might be an overly lenient basis for concluding validity generalization. For example, Kemery, Mossholder, and Dunlap (1989) investigated the use of the 90 percent CV that Pearlman, Schmidt and Hunter (1980) advocated and that, as noted above, has become a standard practice by other Industrial/Organizational psychologists. They concluded that "...the 90% CV consistently provided evidence of transportability when approximately 30 percent of the correlations came from a *rho* of zero..." (Kemery et al., 1989, p. 170).
227. In other words, Kemery et al. (1989) found that the typical 90 percent CV approach to determining evidence of validity generalization actually may lead 30 percent of the time to concluding that validity does generalize when the true validity is no different than zero.
228. In effect, the Barrett et al. (1999) study failed to support validity generalization for cognitive ability tests in predicting the job performance of firefighters, even though the standard Barrett et al. applied was even more lenient than one might think without reviewing the Kemery et al. (1989) research.
229. In our professional opinion, the Barrett et al. (1999) meta-analysis study cited by Drs. Bobko and Schemmer in their expert report does not provide evidence that Written Exams 7029 and 2043 are valid (job related) based on a conclusion of "validity generalization" from

criterion-related validation studies done for other cognitive firefighter tests.

230. After considering the issues discussed above and raised in his deposition, Dr. Bobko's opinion of the Barrett et al. study appears to be less than strong, as well. During his deposition:

Q: In your opinion, what is the quality of this Barrett article?

A: It was higher before an hour ago. (Bobko Deposition, p. 214.)

231. Further, the reference in the report of Drs. Bobko and Schemmer to the earlier Landy et al. (1992) study as "invoking validity generalization" is unsubstantiated. In fact, both admitted during their depositions that they undertook no work to compare the test development plan followed by Landy et al. (1992) with that followed by the City in developing Written Exam 7029 and 2043. (See Bobko Deposition, p. 31; Schemmer Deposition, p. 156).

232. Yet, Drs. Bobko and Schemmer (2007) report that Landy et al. (1992) "invoked validity generalization" from results of a Washington D.C. criterion-related validation study, also executed by Landy Jacobs and Associates, in developing their items for New York City. (Bobko & Schemmer, 2008, p. 27, footnote 34).

233. Drs. Bobko and Schemmer neither present nor discuss any of the earlier study findings on which this "invocation" might be based. Indeed, no document produced by the City provides any information regarding the Landy et al. Washington, D.C. criterion-related validity study.

234. On further exploration, though, Dr. Bobko appears to attach little weight to this "invocation" of validity generalization. In his deposition:

Q: What's your opinion of the transportability study that Landy, Jacobs did as reported in this document?

A: I know there was a transportability study, or I believe there was some use of transportability from one city to another, but I haven't formed an opinion about that aspect of it. I wasn't asked to.

Q: So do you have an opinion as to whether it was a sufficient basis to establish the validity of the written and physical exam established by Landy for New York?

A: I don't have an opinion. (Bobko Deposition, p. 31 - 32).

235. In addition, it is unclear how a single criterion-related validation study such as that of Landy et al. (1992) can be used to posit "validity generalization," a concept typically based upon the meta-analysis of many independent validation studies, as discussed above. The techniques of meta-analysis do not apply to a single study.
236. Landy et al. (1992) might have argued "validity transportability" for the Washington, D.C. criterion-related validation study findings, had they "transported" the same test developed in Washington, D.C. to the City of New York. They did not do this. Instead, they relied upon a showing that the tasks performed by FDNY firefighters were similar to those performed by firefighters in Washington, D.C. to draw two inferences:
- Firefighter ability requirements in the two locations are the same.
  - A new test, designed for the City to measure the same cognitive abilities measured in Washington, D.C., would show the same validity findings noted in Washington, D.C.
237. It is impossible to evaluate these inferences, as the City has produced no information regarding the Landy et al. criterion-related validity study executed in Washington, D.C.
238. The fact that neither of the written examinations now at issue, Written Exams 7029 and 2043, is the examination developed by Landy et al. for Washington, D.C. – or even the one developed by Landy for the FDNY – obviates any argument of validity transportability. The argument is without merit.
239. In our professional opinion, by alluding to the Landy et al. criterion-related validity study conducted in Washington, D.C. as "invoking validity generalization," Drs. Bobko and Schemmer are grasping at straws, reaching for evidence of job-relatedness in an area not supported by any of the information provided by the City.

**E. The cut-off score used to determine pass versus fail status on Written Exam 7029 bears no demonstrated relationship to identifying those candidates minimally qualified to perform the firefighter job.**

- XXIII. The City's job analysis and test development materials provide no explanation of the process followed to arrive at the qualifying (i.e., cut-off) score for Written Exam 7029 – 84.705. No information is provided regarding this score's relationship with expected job performance, its linkage with any aspect of the job analysis process, or its statistical**

connection with factors such as test reliability, standard error of measurement, etc.

240. Before addressing the cut-off score issue, we emphasize again that our discussion assumes (contrary to our actual opinion, detailed above) that Written Exams 7029 and 2043 are valid. Because we conclude that the written examinations are not valid, the test should not be used at all and no qualifying, or cut-off, score would be appropriate.
241. We also note again that, as explained earlier, validity does not reside only in a selection procedure itself, but also in how the scores produced by the procedure are used. As one of the City's experts stated in deposition:
- “[B]ecause validity is as much a function of the process and application of the test as [of] the test itself, . . . certainly validity can be affected by an improper application or inference as validity doesn't follow the test around.” (Schemmer Deposition, p. 244).
242. For example, as Dr. Schemmer admitted during his deposition, no matter how valid the exam, it is the cut-off score that determines who passes and fails a given examination, and a cutoff score unrelated to job performance may well lead to the rejection of applicants who are fully capable of performing the job. (Schemmer Deposition, p. 244).
243. The City does not contend that the cut-off score corresponds to the level necessary to ensure the minimum level of skill needed to perform the job. (See, for example, Defendant's Supplemental Response and Objections to Plaintiff United States' Interrogatories (Nos. 29 – 31), Responses to Interrogatory Nos. 29 and 30).
244. In our opinion, even if we were to assume that Written Exam 7029 were valid, the cut-off score of 84.705 would be above the minimum qualifications necessary to perform the job and thus excluded candidates likely to be successful job performers. During deposition, current and former City representatives agreed. (See, for example, Wachter Deposition, p. 85 – 86 and 179 – 180; Patitucci Deposition, p. 41 – 44).
245. The City, in setting the cut-off score on Written Exam 7029, did not follow professional practice. The result was a cut-off score unrelated to actual performance requirements.
246. In their 1988 publication, Drs. Wayne Cascio, Ralph Alexander, and Gerald Barrett describe methods by which pass versus fail qualifying, or cut-off, scores are set appropriately. The authors describe norm-referenced methods, content-related methods, and criterion-related methods (Cascio, Alexander, & Barrett, 1988).

259. The same criticisms presented above apply with respect to the City's establishment of its qualifying standard for Written Exam 2043.

260. Again, as shown by plaintiff expert Dr. Siskin (2007), this cut-off score resulted in a statistically significant disparate impact upon African-American and Hispanic test takers. Hence, a cut-off score with no demonstrated, or even alleged, relationship with job performance was used in a manner that produced disparate impact.

261. In our professional opinion, the qualifying standard established for Written Exam 2043 is not job-related or consistent with business necessity.

**F. The City's rank-order processing/selection of candidates based on the City's combination of written examination and physical performance test (PPT) scores fails to satisfy either job-relatedness or business necessity.**

XXV. The City combined candidate scores on Written Exam 7029 and the PPT by: a) computing the candidate's standardized written exam score; b) computing the candidate's standardized PPT score; and c) adding the weighted standard scores, each multiplied by .50. This score was then multiplied by 18.472906403940886699, and the value 83.74384236453 was added to produce a final score.<sup>1</sup> There is no evidence that the resulting score is related to the performance of a New York City firefighter.

262. As explained earlier, validity does not reside only in a selection procedure itself, but also in how the scores produced by the procedure are used. Thus, as Dr. Schemmer admitted during his deposition, when composite scores are used, the supporting evidence must pertain directly to the specific score or score combination used. (Schemmer Deposition, p. 244 - 247).

263. For Examination No. 7029, candidates were processed and selected in rank order of their combined written examination and PPT scores (plus bonus points). Based upon a variety of flaws discussed below, it is our professional opinion that use of the rank-ordered list of candidates was not justified.

264. The City offers no justification for the procedure used to compute overall examination scores. In fact, the report of the City's experts, Bobko & Schemmer (2008), offers no support for the process followed

---

<sup>1</sup> Applicable bonus points (i.e., Residency Credit, Legacy Credit, and Veterans Preference points) were added to arrive at a final score for the candidate.



in combining the written and PPT scores, other than to acknowledge that it takes place:

"Exam scores for the written exam and the physical exam were then standardized, and a composite was formed where each of these two components was weighted 50%." (Bobko & Schemmer, 2008, p. 28, footnote 36).

265. During his deposition, Dr. Bobko admitted having provided no opinion regarding the method used to combine candidate scores on the written exam and PPT:

Q: And you don't say anything in your report about whether the method the City used to combine the written exam scores and PPT scores was appropriate, correct?

A: Correct. (Bobko Deposition, p. 329).

266. However, the City's most recent test developer, Dr. Cline, in a May 19, 2005 memorandum to file, stated:

"It seems to me looking at the score distributions, that there is really no "right" way to combine the scores." (Cline Memorandum, 2005, p. 1).

"... I am guessing that something funny may be going on in the physical test score distribution." (Cline Memorandum, 2005, p. 2).

"I think we really need to look at the physical score distribution to see how "abnormal" it is." (Cline Memorandum, 2005, p. 4).

267. We believe Dr. Cline was correct in her concern about the score distribution for the PPT. The City standardized the PPT scores before they were combined with standardized scores on the written examination to generate a rank-ordered list of candidates.

268. The process of standardizing scores (computing z-scores) is based on an assumption of a normal (bell-shaped) distribution of scores. The PPT score distribution was strikingly non-normal. Instead of a smooth bell-shaped curve, a full 69.5 percent of the candidates who passed the PPT and were listed on the final eligibility list received a perfect score on the PPT.

269. This represents a major violation of the assumptions underlying the use of standardized scores. As Dr. Cline admits:

"The z transformation assumes some sort of quasi-normality..."  
(Cline, 2005, p. 1).

270. During his deposition, the City's expert Dr. Schemmer agreed that the distribution of PPT scores was non-normal. He indicated that scores on the written examination were non-normal, as well. (Schemmer Deposition, p. 215 - 218).

**XXVI.** Because the distribution of non-standardized PPT scores is strikingly different from a normal bell-shaped distribution, the resulting standard scores are not useful for rank-ordering candidates. Similarly, when standard scores for the PPT and standard scores for the written examination are combined, the overall scores are, again, not useful for rank-ordering candidates.

271. The City's PPT consisted of eight parts or "events," each of which asked candidates to perform a specific, physically-focused activity that was scored pass-fail, based upon the amount of time the candidate spent performing the event. No differentiation was made among candidates based upon the actual amount of time required for them to complete each event; only upon their ability to complete the event within the allotted time.

272. The City required that candidates pass six of the exam's eight events in order to be eligible for placement on the final rank-ordered eligibility list. As a result, only three possible "raw" passing scores existed for candidates whose PPT results would be combined with their written exam scores: 75 percent, 87.5 percent, and 100 percent, corresponding to passing six, seven, or all eight events. The City provides no justification for this scoring procedure. The City's experts do not even mention it in their report.

273. As Dr. Siskin points out in his July 2008 report, the manner in which the City scored the PPT means that extremely small differences in performance on the PPT could produce unjustifiably large differences in candidates' PPT scores; e.g., failing one event by just one second costs a candidate 12.5 points on the zero to 100 scale used to award PPT points. (Siskin Report, July 2008, p. 22 - 25).

274. Further, a candidate who passed seven of the eight PPT events with exceptional performance on the seven events, but failed the cut-off score on one event by only one second, still would receive a score of 87.5. Yet, another candidate who barely passed all 8 events would receive a score of 100. This aspect of the individual PPT event cut-off scores results in a person who might be less physically capable receiving a higher score than one who is more physically capable, thereby producing



a significant reversal in their final rank order due to the crude and imprecise scoring method applied to the PPT.

275. Dr. Cline, in her memo to file dated May 19, 2005, also commented on the lack of precision in the PPT scores. She stated:

"...it is apparent to me that the physical test is on a much coarser (or cruder) scale than the written test, and does not differentiate among candidates as well as the written test does." (Cline Memorandum, 2005, p. 1).

276. Dr. Cline went on to say:

"Thus, a score of 8 (100%) probably includes people who barely scraped through all eight events, as well as those people who performed really well on all eight events." (Cline Memorandum, 2005, p. 1).

277. In our professional opinion, the City's method of scoring the PPT is not useful for differentiating between and among candidates. Differences in candidates' physical abilities can be exaggerated in many instances, and masked in others. Any rank ordering of candidates that is based, even in part, on such a scoring method is flawed.

278. Moreover, the City has not justified the cut-off scores/times on each of the eight PPT events. The report of the City's experts, Drs. Bobko and Schemmer, does not mention the PPT cut-offs. Notably, the physical test the City adopted for use as part of Examination No. 6019 (the CPAT), which consists of eight events much like those included in the PPT, does not have individual event cut-offs and is scored based on the candidate's overall performance on all eight events.

279. According to the report of Dr. Landy (1999), the expert who originally implemented the PPT for the City in 1992, the:

"...cut scores for each event were calculated at a raw score corresponding to the 16<sup>th</sup> percentile of incumbent performance for each event." (Landy, 1999, p. 15).

280. However, the incumbent firefighters Landy (1999) used to establish these cut-off scores for each event were both probationary firefighters who had just completed academy training, and non-probationary firefighters. Landy (1999) did not report the number of probationary versus non-probationary firefighters who were in the sample, nor did he report the percent of the sample that consisted of probationary firefighters.

281. Dr. Landy (1999) acknowledged that:

“[n]ormally, the incumbent sample would only include a representative sample of the fire department’s working firefighters. However, due to a lack of union support, the participation of probationary firefighters was needed as well.” (Landy, 1999, p. 14).

The result is that the cut-off score for each of the eight events would be expected to fail 16 percent of newly hired firefighters who, as a group, were very likely to perform better on the PPT events than is typical for a “representative sample of the fire department’s working firefighters.”

282. Hence, even if the standardization process were a reasonable procedure for the data set (which it was not), the raw, non-standardized PPT score with which the process commences lacks justification.

283. Written Exam 7029 also has imprecision in its measurement of cognitive abilities. As described in earlier sections of our report, Written Exam 7029 does not include measures of many skills and abilities important to effective performance in the job of firefighter. As a result, the final rank ordering of candidates on Examination No. 7029 is devoid of important information about candidate skills and abilities that are required in order to perform the job.

**XXVII. The illusion of precision in computing candidates’ scores to more than ten decimal places is contradicted by the fact that the City does not even posit an estimate of the reliability associated with candidates’ combined written examination and PPT scores. In fact, the reliability of the combined score appears to be unknown.**

284. In spite of all the problems discussed above, and the resulting lack of accuracy in measuring needed skills and abilities, the City, in calculating each candidate’s final score for Examination No. 7029, multiplied each candidate’s combined score on the written examination and PPT by 18.472906403940886699 and then added 83.74384236453.

285. The resulting appearance of accuracy is an illusion. In truth, candidates on the final eligibility list received one of only 42 different possible scores; one of only 14 possible passing scores on the written exam (72 through 85 items answered correctly) and one of only three possible passing scores on the PPT (six, seven, or eight exercises passed).

286. Although the illusion is less egregious for Examination No. 2043, the numbers (i.e., 12.7226 and 88.4606) still suggest a much greater accuracy of measurement than is present.